RESIDUAL ENTROPY FUNCTION AND ITS APPLICATIONS

DISSERTATION

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD 0F THE DEGREE OF

MASTER OF PHILOSOPHY

IN

STATISTICS

BY

NUSRAT MUSHTAQ

UNDER THE SUPERVISION OF

DR. M. A. K. BAIG

ASSOCIATE PROFESSSOR

POST GRADUATE DEPARTMENT OF STATISTICS
UNIVERSITY OF KASHMIR, Srinagar
(NAAC ACCREDIATED GRADE ‘A’)
(2012)

DEDICATED

TO

MY PARENTS

Acknowledgement

"Dreams become reality when
the efforts you put in are sincere"

First and foremost I bow down before Almighty "ALLAH" with solicit and imploreness to express my deepest sense of gratitude for his blessings, mercy, faithfulness and thus enabling me to accomplish this venture. No human effort could climb the ladder without being given right direction.

It is my pleasant privilege to acknowledge with gratitude one and all from whom I received guidance, cooperation and inspiration during the tiresome trials of this study.

I wish to express my deep sense of gratitude and indebtedness to Dr. M. A. K. Baig, Associate Professor, Department of Statistics, university of Kashmir, Srinagar for his enthusiastic guidance, kind supervision, continued encouragement, generous assistance, timely cooperation, valuable suggestions and constructive criticism during the course of present study.

Thanks are due to Professor Aquil Ahmad, Head, Department of Statistics, University of Kashmir, Srinagar for his moral support, useful suggestions and continuous encouragement all through the course of study and providing the necessary facilities in the department. My words fail to express sincere gratitude to him.

I also express my profound gratitude to my teachers Dr. Tariq Rashid Jan, Dr. Anwar Hassan and Dr. S. Parvaiz Ahmad for their cooperation, whole hearted support and timely help.

I gratefully acknowledge the encouragement and moral boosting from Mr. Irfan.

I would like to put on record my thanks to the technical staff (especially Wani Vaseem) and non-teaching staff members of Statistics Department University of Kashmir for their cooperation and timely help.

The successful completion of the research programme was possible only because of the active cooperation of my friends and colleagues. I am particularly indebted to Ms. Humaira Sultan, Ms. Saba Riyaz, Ms. Suriya Jabeen, Mr. Mohd Javid, Mr. Raja Sultan, Mr. Adil Rashid, Mr. Adil Hamid, Mr. Shabir, Mr. Irfan and Mr. Kaisar for their help and close cooperation.

I am highly indebted to my parents who have been the major source of inspiration behind my present endeavor as their well wishes and prayers always escorted me to the completion of this programme.

I take this opportunity to record my deep sense of appreciation for my sisters Shaheen, Masrat, Rifat, Ushrat, and Rutbaa and my brothers Raman and Manzoor and my nephew (Azaan) and niece (Aleeza) and brother-in-law (Irshad Ahmad) who were always a source of inspiration and shouldered all the domestic responsibilities during the course of my research.

Lastly I thank Mr. Parvaiz Ahmad for printing the dissertation with patience.

I owe the responsibility for any error, omission, typographical mistakes and other things alike that have incorporated in this manuscript.

Nusrat Mushtaq

Preface

The concept of information originated when an attempt was made to create a theoretical model for the transmission of information of various kinds. Information theory is a branch of Mathematical theory of probability and is applied in a wide variety of fields: Communication Theory, Thermodynamics, Economics, Cybernetics, Operation Research and Psychology.

Much work has been done on this branch of probability and it has acquired a great currency in various research Journals of Statistics. In this light, I compiled my dissertation on the topic ***"Residual Entropy Function and Its Applications"*** and the chapter wise scheme is as follows:

Chapter - I gives the basic concepts and preliminary results which are used in subsequent chapters. This makes the rest of the dissertation readable.

Chapter - II the concept of residual entropy have been introduced which measures the uncertainty of a component that has survived for some unit of time. Characterization results of some lifetime models have been studied by developing a functional relationship between the residual entropy function and hazard rate function and their properties are discussed.

Chapter - III deals with the past residual entropy and its characterization results. Some ordering and aging properties have been defined in terms of past entropy and their properties have been discussed. A non parametric class based on generalised past entropy and its properties are also discussed.

Chapter - IV throws light on some new results on cumulative residual entropy and its properties. Some characterization results and conditional cumulative residual entropy have also been discussed.

Chapter - V discusses the generalized residual entropy of order statistics and record values. Its characterization results and applications are also studied.

The intent of this manuscript is to present a survey of the existing literature on residual entropy function and its applications. It will be a useful document for the future researchers in this area. The area of residual entropy function and its applications is fertile and there is a lot of scope to work on this concept.

Nusrat Mushtaq

CONTENTS

POST GRADUATE DEPARTMENT OF STATISTICS
UNIVERSITY OF KASHMIR, SRINAGAR

No:_____________
Dated:___________

Certificate

This is to certify that the scholar **Nusrat Mushtaq** *has carried out the present dissertation entitled* ***"RESIDUAL ENTROPY FUNCTION AND ITS APPLICATIONS"*** *under my supervision and the work is suitable submission for the award of the Degree of Master of Philosophy in Statistics. It is further certified that the work has not been submitted in part or full for the award of M. Phil or any other degree.*

Dr. M. A. K. Baig
(Supervisor)

Chapter - I

Basic Concepts and Preliminary Results

INTRODUCTION

The word ‘information’ is very common word used in everyday language. Information transmission usually occurs through human voice (as in telephone, radio, television, etc.), books, newspapers, letters, etc. In all these cases a piece of information is transmitted from one place to another. However, one might like to quantitatively assess the quality of information contained in a piece of information. Few examples are as follows:

1. Suppose, one states, ‘It is raining’. Now the question is ‘Have we received much information?’ Here it may be concluded that if a piece of information is presented, which was already known, then, obviously, no information has been received. Again, if one states that ‘The sun will shine the whole day tomorrow’. In this case an information has been received without being specific. Since we have been informed that something will happen about which we did not know, therefore, we do not have to be much surprised by the statement that was made.
2. Suppose, we have come to know from the weather forecast on television that ‘The rain will continue for the next two days’. In this case, we have received information more than in the first example above because a statement has been made whose truth is not at all so surprising.

Information theory is a new branch of probability theory with extensive potential applications to the communication systems. Like several other branches of mathematics, information theory has a physical origin. It was principally originated by C.E. Shannon [1948], through two outstanding contributions to the mathematical theory of communications. These were followed by a flood of research papers speculating upon the possible applications of the newly born theory to broad spectrum of research areas such as pure mathematics, psychology, economics, biology, etc.

The first attempt to develop the mathematical measure for communication channels was made by Nyquise {[1924], [1928]} and Hartley [1928]. The main contributions which really gave birth to the so called information theory, came shortly after the second world war from the mathematicians C. E. Shannon [1948] and N. Wiener [1961]. In the paper entitled, "The Mathematical Theory of Communication" Shannon made the first attempt to deal with the new concept of the amount of information and its main consequences. Perhaps the most important theoretical result of information theory is the Shannon's fundamental theorem in which he first set up a mathematical model for quantitative measure of average amount of information provided by a probabilistic experiment and proved a number of interesting results which showed the importance and usefulness measure of information.

In the last 40 years, the information theory has been more precise and has grown into staggering literature. Some of its terminology even has become part of our daily language and has been brought to a point where it has found its wide applications in various fields of importance. e.g., The work of Bar-Hillel [1964], B. Subrahmanyan and Siromoney [1968] in Linguistic, Brillouins [1956] in physics, Theil [1967] in economics, Quastler [1954] in Psychology, Quastler [1953] in Biology and Chemistry, Wiener [1961] in Cybernetics, Kerridge [1961] in Statistical estimation, Kapur [1968] in

Operation Research, Kullback [1959] in Mathematical Statistics, Zaheerudin [1987] in Inference, Zadeh [1996] in Fuzzy set theory, Ebrahimi [1996] in Survival analysis, Rao [1982] in Anthropology, Mei [1978] in Genetics, Sen [1973] in Political Science and Chen [1973] in pattern recognition. Jaynes [1975] first stated the maximum entropy principle explicitly and during last three decades, the principle has been applied with the varying degree of success in fields such as Thermodynamics and Statistical Mechanics, Design of Experiment and Contingency Tables, Search theory, Reliability theory, Banking, Insurance, Accountancy and Marketing, Transportation problems, etc. We restrict ourselves only to those aspects of information theory which are closely related to our research work.

In the context of reliability and lifetime distributions, there are some measures such as the hazard rate function or the mean residual lifetime function that have been used to characterize or compare the aging process of a component. Cox [1972] and Kotz and Shanbhag [1980] have shown that both the functions determines the distribution function uniquely. Ebrahimi [1996] proposed an alternative characterization of a lifetime distribution in terms of conditional Shannon's entropy. Based on the measure of residual entropy, Ebrahimi and Pellery [1995] and Ebrahimi and Kirmani [1996] have studied some ordering and aging properties of lifetime distributions. Belzunce et al. [2004] extended some results given by Nair and Rajesh [1998] and Asadi and Ebrahimi [2000] to characterize a distribution from functional relationships between the residual entropy and the mean residual life or hazard rate function. Various generalizations of Ebrahimi's measure have been proposed by many researchers including Abraham and Sankaran [2006], Nanda and Paul [2006] and Hooda and Kumar [2007]. Measure of uncertainty in past lifetime distributions have been proposed by Crescenzo and Longobadri [2002] and generalized by Nanda and Paul [2006]. Ebrahimi and Kirmani [1996] and Gupta and Nanda [2002] gave an overview of some aspects of residual divergence measures and studied some characterization theorems under the assumption that the distribution function satisfy the Cox's proportional hazard rate model.

1.1 Information Function and Shannon's Entropy

1.1.1 Information Function: Let E_i be the ith event with probability of occurrence p_i , the information function may be defined as

$$h(p_i) = -\,log(p_i) \qquad (1.1.1)$$

1.1.2 Shannon's Entropy: Let S be the sample space belongs to random events. Compose this sample space into a finite number of mutually exclusive events $E_1, E_2, \dots, E_n$, whose respective probabilities are $p_{1,}\, p_2, \dots p_n$, then the average amount of information or Shannon's entropy is defined as

$$H(P) = E\big(h(p_i)\big) = -\sum_{i=1}^{n} p_i\, log\, p_i\,, \qquad 0 < p_i < 1 \qquad (1.1.2)$$

Some important properties of Shannon's entropy is given below:

(I) Continuity: $H(P)$ is continuous P, i.e. the measure should be continuous, so that changing the values of the probabilities by a very small amount should only change the entropy by a small amount.

(II) Symmetry: The measure remains unchanged if the outcomes x_i are re-ordered.

$$H(p_1, p_2) = H(p_2, p_1).$$

(III) Maximality: The measure should be maximal if all the outcomes are equally likely (uncertainty is highest when all possible events are equiprobable).

$$\max H\,(p_1, \dots, p_n) = H\left(\frac{1}{n}, \dots \frac{1}{n}\right)$$

(IV) Additivity: The additivity property states that for two independent probability distributions $= (p_1, p_2, \dots, p_n), Q = (q_1, q_2, \dots, q_m)$,

$$H(p_1, p_2, \ldots, p_{n-1}, q_1, q_2, \ldots, q_m) =$$

$$H(p_1, p_2, \ldots, p_{n-1}, p_n) + p_n H\left(\frac{q_1}{p_n}, \frac{q_2}{p_n}, \ldots, \frac{q_m}{p_n}\right),$$

where $p_n = \sum_{k=1}^{m} q_k$.

In addition to the above four basic properties, we have the following properties

(v) Expansibility: The value of the entropy function should not change, if an impossible outcome is added to the probability scheme, i.e.

$$H_{m+1}(p_1, p_2, \ldots, p_m, 0) = H_m(p_1, p_2, \ldots, p_m).$$

(vi) For the two independent probability distributions

$$P = (p_1, p_2, \ldots, p_m), \quad Q = (q_1, q_2, \ldots, q_n),$$

where $\sum_{i=1}^{m} p_i = 1$, $\sum_{j=1}^{n} q_j = 1$,

then the uncertainty of the joint scheme should be the sum of their uncertainties, i.e.,

$$H_{mn}(P \cup Q) = H_m(P) + H_n(Q).$$

(vii) Normality: The entropy becomes unity for two equally probable events, i.e.

$$H\left(\frac{1}{2}, \frac{1}{2}\right) = 1.$$

1.1.3 Unit of Information: When the logarithm is taken with the base 2, the unit of information is called bit, when the natural logarithm is taken then the resulting unit is called Nat and if the logarithm is taken with the base 10, the unit of information is known as Hartley.

It must be noted that the definition of Shannon's entropy though defined for a discrete random variable can be extended to situations when the random variable under consideration is continuous.

Let X be a continuous random variable with the density function $f(x)$ on I, where $I = (-\infty, \infty)$, then the entropy is defined as

$$H(X) = -\int_{\infty}^{\infty} f(x) \, log \, f(x) dx \qquad (1.1.3)$$

whenever it exists. The measure (1.1.3) is also called differential entropy. It has many of the properties of discrete random entropy but unlike the entropy of the discrete random variable, the differential entropy may be infinitely large, negative or positive, Ash [1990]. Also, the entropy of the discrete random variable remains invariant under a change of variable. However with a continuous random variable the entropy does not necessarily remains invariant.

1.2 Generalizations of Shannon's Entropy

Various generalizations of Shannon's entropy are available in the literature. Some important generalizations are given below.

(i) Renyi's Entropy: Renyi [1961] generalized the Shannon's entropy by defining the entropy of order α as

$$H_\alpha(P) = \frac{1}{1-\alpha} log \sum_{i=1}^{n} p_i^\alpha \,, \qquad \alpha > 0 (\neq 1) \qquad (1.2.1)$$

and in continuous case

$$H_\alpha(X) = \frac{1}{1-\alpha} log \int_{0}^{\infty} f^\alpha \, (x) dx \,, \quad \alpha > 0 (\neq 1) \qquad (1.2.2)$$

For $\alpha \to 1$, the measure (1.2.1) and (1.2.2) reduces to (1.1.2) and (1.1.3) respectively.

(ii) Havrada- Charavat's Entropy: Havrada- Charavat [1967] introduced the entropy as

$$H^{\beta}(P) = \frac{1}{1-\beta}\left(\sum_{i=1}^{n} p_i^{\beta} - 1\right), \quad \beta > 0(\neq 1) \qquad (1.2.3)$$

and in continuous case

$$H^{\beta}(X) = \frac{1}{1-\beta}\left(\int_0^{\infty} f^{\beta}(x)dx - 1\right), \quad \beta > 0(\neq 1) \qquad (1.2.4)$$

and is called generalized entropy of type β. When $\beta \to 1$, the measure (1.2.3) and (1.2.4) becomes Shannon's measure (1.1.2) and (1.1.3) respectively.

(iii) Varma's Entropy: Varma [1966] introduced the entropies as

$$H_{\alpha}^{\beta}(P) = \frac{1}{\beta-\alpha} log\left(\sum_{i=1}^{n} p_i^{\alpha+\beta-1}\right), \quad \beta - 1 < \alpha < \beta, \beta \geq 1 \qquad (1.2.5)$$

and in continuous case

$$H_{\alpha}^{\beta}(X) = \frac{1}{\beta-\alpha} log \int_0^{\infty} f^{\alpha+\beta-1}(x)dx, \beta - 1 < \alpha < \beta, \beta \geq 1 \qquad (1.2.6)$$

For $\beta = 1, \alpha \to 1$, the measure (1.2.5) and (1.2.6) reduces to (1.1.2) and (1.1.3) respectively.

(iv) Arimoto's entropy: Armito [1971] introduced the generalized entropy as

$$A_{\alpha}(P) = \frac{1}{2^{\alpha-1}-1}\left\{\left(\sum_{i=1}^{n} p_i^{\frac{1}{\alpha}}\right)^{\alpha} - 1\right\}, \quad \alpha > 0(\neq 1) \qquad (1.2.7)$$

and in continuous case

$$A_{\alpha}(X) = \frac{1}{2^{\alpha-1}-1}\left\{\left(\int_0^{\infty} f^{\frac{1}{\alpha}}(x)dx\right)^{\alpha} - 1\right\}, \quad \alpha > 0(\neq 1) \qquad (1.2.8)$$

For $\alpha \to 1$, (1.2.7) and (1.2.8) reduces to (1.1.2) and (1.2.3) respectively.

(v) Boekee and Lubbe's Entropy: Boekee and Lubbe [1980] introduced the generalized entropy as

$$H_R(P) = \frac{R}{R-1}\left\{1 - \left(\sum_{i=1}^{n} p_i^R\right)^{\frac{1}{R}}\right\}, \qquad R > 0(\neq 1) \qquad (1.2.9)$$

and in continuous case

$$H_R(X) = \frac{R}{R-1}\left\{1 - \left(\int_0^{\infty} f^R(x)dx\right)^{\frac{1}{R}}\right\}, \qquad R > 0(\neq 1) \qquad (1.2.10)$$

For $R \to 1$, (1.2.9) and (1.2.10) reduces to Shannon's entropy given in (1.1.2) and (1.1.3) respectively.

(vi) Kapur's Entropy: Kapur $[1967]$ generalized the Shannon's entropy as

$$H_{\alpha,\beta}(P) = \frac{1}{1-\alpha} log \frac{\sum_{i=1}^{n} p_i^{\alpha+\beta-1}}{\sum_{i=}^{n} p_i^{\beta}} \qquad (1.2.11)$$

where $p_i \geq 0,\ \alpha > (\neq 1),\ \beta > 0.$

and in continuous case

$$H_{\alpha,\beta}(X) = \frac{1}{1-\alpha} log \frac{\int_0^{\infty} f^{\alpha+\beta-1} dx}{\int_0^{\infty} f^{\beta}(x)dx}, \quad \alpha > (\neq 1), \beta > 0 \qquad (1.2.12)$$

For $\beta = 1, \alpha \to 1$, the measure (1.2.11) and (1.2.12) reduces to (1.1.2) and (1.1.3) respectively.

(vii) Sharma and Mittal's Entropy: Sharma and Mittal $[1977]$ introduced the generalized entropies as

$$H_\alpha(P) = \frac{1}{(2^{1-\alpha}-1)}\left[\exp\left((\alpha-1)\sum_{i=1}^{n} p_i \, log \, p_i\right) - 1\right],$$

$$\alpha > (\neq 1) \qquad (1.2.13)$$

and in continuous case

$$H_\alpha(X) = \frac{1}{(2^{1-\alpha}-1)}\left[\exp\left((\alpha-1)\int_0^\infty f(x)\, log\, f(x)dx\right)-1\right],$$

$$\alpha > (\neq 1) \qquad (1.2.14)$$

For $\alpha \to 1$, (1.2.13) and (1.2.14) reduces to (1.1.2) and (1.1.3) respectively.

(viii) Sharma and Taneja's Entropy: Sharma and Taneja [1975] introduced the generalized entropies as:

$$H_\alpha(P) = -2^{\alpha-1}\sum_{i=1}^{n} p_i^\alpha \log p_i\,, \qquad \alpha > 0 \qquad (1.2.15)$$

For $\alpha = 1$, (1.2.15) reduces to (1.1.2).

$$H_{\alpha,\beta}(P) = \frac{1}{2^{1-\beta}-2^{1-\alpha}}\sum_{i=1}^{n}\left(p_i^\beta - p_i^\alpha\right), \alpha \neq \beta, \alpha, \beta > 0 \quad (1.2.16)$$

For $\beta = 1$, $\alpha \to 1$, (1.2.16) reduces to Shannon's entropy (1.1.2).

For continuous cases, we have the following generalizations

$$H_\alpha(X) = -2^{\alpha-1}\int_0^\infty f^\alpha\,(x)\, log\, f(x)dx, \qquad \alpha > 0 \qquad (1.2.17)$$

$$H_{\alpha,\beta}(X) = \frac{1}{2^{1-\beta}-2^{1-\alpha}}\int_0^\infty \left(f^\beta(x) - f^\alpha(x)\right)dx \quad \alpha \neq \beta, \alpha, \beta > 0$$

$$(1.2.18)$$

For $\alpha = 1$, (1.2.17) reduces to (1.1.3) and for $\beta = 1$, $\alpha \to 1$, (1.2.18) reduces to Shannon's entropy (1.1.3).

(ix) Kerridge Inaccuracy: Suppose that an experiment asserts that the probabilities of n events are $Q = (q_1, q_2, .., q_n)$ while their true probabilities are $P = (p_1, p_2, \dots, p_m)$, then the Kerridge [1961] has proposed the inaccuracy measure as

$$K(P;Q) = -\sum_{i=1}^{n} p_i \log q_i \tag{1.2.19}$$

When $p_i = q_i \ \forall\ i = 1, 2, \dots, n$, then (1.2.19) reduces to Shannon's entropy.

In case of continuous distribution

$$K(f;g) = \int_0^{\infty} f(x) \log g(x) dx \tag{1.2.20}$$

1.3 Joint and Conditional Entropy

1.3.1 Joint Entropy: The joint entropy $H(X;Y)$ of a pair of discrete random variable $(X;Y)$ with a joint distribution function $p(x;y)$ is defined as

$$H(X;Y) = -\sum_{x \in \mathrm{R}} \sum_{y \in R} p(x;y) \log p(x;y) \tag{1.3.1}$$

and in continuous case

$$H(X;Y) = -\int_0^{\infty} \int_0^{\infty} f(x;y) \log\ f(x;y) dx\, dy \tag{1.3.2}$$

where $f(x;y)$ is the joint density function of the random variable X and Y.

1.3.2 Conditional Entropy: The entropy is meant to measure the uncertainty in the realization of X. Now, we want to quantify how much uncertainty does the realization of a random variable X have if the outcome of another random variable Y is known. This is called conditional entropy and is given by:

$$H(X/Y) = -\sum_{x,y} P(x,y) \log \frac{P(x,y)}{P(y)} \tag{1.3.3}$$

where $p(y)$ is the marginal distribution of Y.

And in continuous case

$$H(X/Y) = -\int_{0}^{\infty} f(x,y)\, log \frac{f(x,y)}{f(y)}\, dx \qquad (1.3.4)$$

1.4 The Survival Analysis

1.4.1 Cumulative Distribution Function: If X is a continuous random variable with the probability density function $f(x)$, then the function

$$F_X(t) = P(X \leq t) = \int_{-\infty}^{t} f(x)dx\,, \qquad -\infty < x < t$$

is called cumulative distribution function of the random variable X. The distribution function has the following properties

(i) $F(t)$ is non-decreasing function in t, i.e.

$$F'(t) = \frac{d}{dt}F(t) = f(t) \geq 0.$$

(ii) $F(-\infty) = 0$ and $F(+\infty) = 1$, which implies that

$$0 \leq F(x) \leq 1.$$

(iii) $F(t)$ is a continuous function of t on the right.

(iv) It may be noted that

$$P(a \leq X \leq b) = F(b) - F(a).$$

Similarly

$$P(a < X < b) = \int_{a}^{b} f(t)dt.$$

1.4.2 Survival Function: The basic quantity employed to describe time-to-event phenomena is the survival function. This function, also known as reliability function is the probability that an individual survives beyond time t.

If X is a continuous random variable then the survival function which is usually denoted by $R(t)$, is defined by

$$R(t) = P(X \geq t) = \int_t^\infty f(x)dx \qquad (1.4.1)$$

In the context of equipment or manufactured item failures, $R(t)$ is referred to as the reliability function. Note that the survival function is a non-increasing function with $R(0) = 1$ and $R(\infty) = 0$.

Thus, we have the following relationship between reliability function and distribution function

$$R(t) = 1 - F(t) \qquad (1.4.2)$$

Differentiating (1.4.2) both sides with respect to t, we have

$$\frac{d}{dt}R(t) = -f(t)$$

or

$$f(t) = -\frac{d}{dt}R(t) \qquad (1.4.3)$$

1.4.3 The Hazard Rate Function: It is the probability that the item will fail in the next δt time unit given that the item is functioning properly in time t. In other words, failure rate or hazard rate function is defined as the conditional probability of failure between $(t, t + \delta t)$ given that there is no failure up to time t,

$$\lambda(t) = \lim_{\delta t \to 0} P[t \leq T \leq t + \delta t / T > t] \qquad (1.4.4)$$

or

$$\lambda(t) = \frac{f(t)}{\bar{F}(t)} \qquad (1.4.5)$$

If $F(t)$ be the distribution of time to failure and $f(t)$ be the p.d.f, then we have

$$R(t) = 1 - F(t) \tag{1.4.6}$$

Therefore (1.4.5) reduces to

$$\lambda(t) = \frac{f(t)}{1 - F(t)} \tag{1.4.7}$$

1.4.4 Distribution with Increasing Failure Rate (IFR) and Decreasing Failure Rate (DFR): Its often difficult to single out a specific model to characterized behavior of a system or a device consequently a less conventional approach. Where in the failure behavior is characterized merely by property of hazard rate (failure rate) is often to found to be quite useful. Such an approach has not only alleviated the task of specifying failure model but also initiated the develop of comprehensive theory of reliability. The measure of an equipment reliability in frequency with which failure occur in time. A failure distribution represents an attempt to discrete mathematically, the length of the life of the material or a device, there are many physical causes that individual or collectively may be responsible for the failure of the device at any particular instance, the hazard function describes the way in which the instance probability of death individual change with time. Let $F(t)$ be the distribution of time to failure and $f(t)$ be the p.d.f. then hazard rate is defined as

$$\lambda(t) = \frac{f(t)}{\bar{F}(t)} \tag{1.4.8}$$

We have $\bar{F}(t) = 1 - F(t)$, then

$$\lambda(t) = \frac{f(t)}{1 - F(t)} \tag{1.4.9}$$

$F(t)$ is called survival function denoted as $\bar{F}(t)$. The failure rate which is the function of time has probabilistic interpretation namely $\lambda(t)dt$ represents the probability that a device of age t will fail in interval $[t + (t + \delta t)]$ or $\lambda(t) =$

p{a device of age t will fail in interval $(t, t + \delta t)$ / device of function in time t}. Now in considering a life testing model, its often more informative to consider the properties of hazard function then characterized the model, in terms of p.d.f or c.d.f. directly the monotonically hazard rate is an important consideration. If $\lambda(t)$ is Hazard Function (HF), such that

$$t_1 \leq t_2, \quad \lambda(t_1) = \lambda(t_2) \tag{1.4.10}$$

The considering model is said to be increasing failure rate (IFR).

If $\lambda(t)$ is Failure Function, such that

$$t_1 \geq t_2, \quad \lambda(t_1) \geq \lambda(t_2) \tag{1.4.11}$$

The corresponding model is said to be decreasing failure rate (DFR).

A decreasing failure rate might be interpreted an improvement of unit with age.

Several alternatives criteria for assisting whether F is (IFR) DFR distribution exists.

1. F is increasing (decreasing) failure rate distribution if $F(t + \delta t) - F(t)/1 - F(t)$ is increasing (decreasing) in time t for all $\delta t > 0$.
2. F is increasing (decreasing) failure rate, if $log[(1 - F(t))]$ is concave (convex) for all $\delta t > 0$.

1.4.5 Average Failure Rate (AFR): The average failure rate is defined in terms of the function

$$A(x) = \frac{1}{x}\left[\int_0^x h(t)dt\right] \tag{1.4.12}$$

A life testing model is said to have an increasing failure rate average (IFRA) if $x_1 \leq x_2$ implies $A(x_1) \geq A(x_2)$. On the other hand, a life testing model is said to have a decreasing failure rate average (DFRA) if $x_1 \leq x_2$ implies $(x_1) \geq A(x_2)$.

1.4.6 The Mean Residual Life: The mean residual life (MRL) is denoted by $\mu(x)$ and is defined by

$$\mu(x) = \frac{\int_x^\infty (t-x)\, f(t)dt}{1-F(x)} \tag{1.4.13}$$

The MRL is the generalization of the mean life of a unit, since

$$\mu(0) = \int_x^\infty t\, f(t)dt \;= E(x) \tag{1.4.14}$$

One possible interpretation of the MRL involves the conditional distribution of X given $X > x$ in particular for a fixed, consider the $f(t/X > x) = \frac{f(t)}{1-F(t)}$, function, if $t > x$ and zero otherwise the function $f(t/X > x)$ is the conditional p.d.f. of x given $X > x$ and consequently. The MRL is the conditional expectation of $X - x$ given $X > x,\ \mu(x) = E(X/X > x) - x$. In other words MRL is the average amount of unused life of unit at age x. A lifetime model is said to have a decreasing mean residual life (DMRL), if $x_1 \leq x_2$ implies $\mu(x_1) \geq \mu(x_2)$ .On the other hand a life-testing model is said to have sn increasing mean residual life (IMRL), if $x_1 \leq x_2$ implies $(x_1) \geq \mu(x_2)$.

1.4.7 Some Characterization Results

We have from (1.4.1)

$$R(t) = \exp\left\{-\int_0^t h(x)dx\right\}$$

$$= \exp\{S(t)\}.$$

Therefore

$$R(t) \ = \frac{r(0)}{r(t)} \exp\left\{-\int_0^t \frac{1}{r(x)} \ dx\right\} \tag{1.4.15}$$

But,

$$f(t) = -\frac{d}{dt}R(t)$$

$$= h(t)R(t).$$

Thus

$$f(t) = \left(\frac{d}{dt}r(t) + 1\right)\frac{r(0)}{r^2(t)} \exp\left\{-\int_0^t \frac{1}{r(x)} \mathrm{d}x\right\} \tag{1.4.16}$$

Dividing (1.4.15) to (1.4.16), we get

$$\frac{f(t)}{R(t)} = \frac{\left(\frac{d}{dt}r(t) + 1\right)}{r(t)}$$

Thus

$$h(t) = \frac{r'(t) + 1}{r(t)} \tag{1.4.17}$$

where $r'(t) = \frac{d}{dt}r(t)$. Equation (1.4.11) gives the functional relationship between hazard rate function and mean residual life function. It has a pivot role in some characterization results of lifetime models by using the information theoretic approach.

1.4.8 Reversed Hazard Rate Function

The concept of reversed hazard rate was initially introduced as the hazard rate in the negative direction and received the cold reception in the literature at the early stage. This was because reversed hazard rate, being the ratio of probability density function and the corresponding distribution function, was conceived as a dual measure of hazard rate.

Keilson and Sumita [1982] were among the first to define reversed hazard rate and called it the dual failure function. According to them, hazard rate ordering is the uniform stochastic ordering and the reversed hazard rate ordering is the uniform stochastic ordering in the negative direction. This has followed by Shaked and Shanthikumar [1994], who have presented some nice results relating to reversed hazard rate function. Also, what is important is the inclusion of some interesting characterizations based on the monotonicity of reversed hazard rate function.

Let X be a continuous random variable with density function $f(x)$, cumulative distribution function $F(x)$ and survival function $R(x)$. Then, the reversed hazard rate of X at t is denoted by $\tau(t)$ and is defined as

$$\tau(t) = \frac{d}{dt} log\, F(t) = \frac{f(t)}{F(t)}$$

The following relationship can be easily obtained

$$F(t) = \exp\left\{-\int_{t}^{\infty} \tau(x)dx\right\}.$$

1.4.9 Discrete Case

Let X is a discrete random variable taking the values $x_1 < x_2 < \cdots < x_n$ with the probability mass function, $P(j) = P(X = j)$, $j = 1,2, \dots, n$ then the survival function is defined as

$$R(j) = \sum_{k=j}^{n} P(k) \qquad (1.4.18)$$

The survival function and the probability mass function are related by

$$P(j) = R(j) - R(j+1) \qquad (1.4.19)$$

The hazard function is defined as

$$h(j) = \frac{P(j)}{R(j)} \qquad (1.4.20)$$

Using (1.4.20), we have

$$h(j) = 1 - \frac{R(j+1)}{R(j)} \qquad (1.4.21)$$

The survival function is related to the hazard rate function by

$$R(j) = \prod_{k=j}^{n} [1 - h(k)] \qquad (1.4.22)$$

For discrete lifetimes the cumulative hazard function is defined as

$$S(j) = \sum_{k=j}^{n} h(k) \qquad (1.4.23)$$

The characterization relationship between survival function, hazard rate function and mean residual lifetime can be developed as:

We have from (1.4.18)

$$R(j) = \sum_{k=j}^{n} P(k)$$

$$= \prod_{k=j}^{n} [1 - h(k)].$$

If X is an integer valued random variable with mean residual life at time k equal to $m_k, k = 0, 1, 2, \ldots$ and m_0 is finite then we have

$$R(k) = \frac{1 + m_0}{m_k} \prod_{j=0}^{k} \frac{m_j}{1 + m_j} \qquad (1.4.24)$$

Also, for any discrete survival function, we have

$$P(j) = R(j) - R(j+1)$$
$$= h(j)R(j).$$

Therefore,

$$h(j) = \frac{P(j)}{R(j)}.$$

1.5 Classes of Aging Distributions

An important characteristic of survival distribution is its aging properties. There are number of classes that have been suggested in the literature to categorize distributions based on their aging properties or their dual. The first class is the class of increasing hazard rate *(IHR)* distributions and the dual class of decreasing hazard rate *(DHR)* distributions. A Survival distribution is said to be in the *IHR (DHR)* class if and only if $h(t)$ is increasing (decreasing) for all t.

A second more general aging class is the class of increasing (decreasing) hazard rate on the average, *IHRA (DHRA)* distributions. A distribution is said to fall in the *IHRA (DHRA)* class if and only if

$$-\left(\frac{1}{t}\right)\log[R(t)] \qquad (1.5.1)$$

is increasing (decreasing) in t.

The definition arises by declaring a distribution to be in the *IHRA* class when its cumulative hazard rate, $-\log[S(t)]$ is increasing faster than the cumulative hazard rate of an exponential random variable. Since the exponential distribution reflects a model with no aging, this class is one of distributions for which individuals are on the average aging.

Since (1.5.1) implies that $R^{\frac{1}{t}}(t)$ is increasing in t, we have that X is in *IHRA* class if and only if $R(\theta t) \geq R^{\theta}(t)$.

A third aging class is the class of decreasing (increasing) mean residual life, *DMRL (IMRL)* distributions. A distribution is said to be *DMRL (IMRL)* class if

$$r(t) = \frac{\int_t^{\infty} R(x)dx}{R(t)} \tag{1.5.2}$$

is increasing (decreasing) in t.

This aging class, which include all *IHR* models, is one where the mean remaining life of an individual of age t is becoming shorter as t increases.

1.6 Some Mathematical Functions

1.6.1 Convex function: A real valued function $f(x)$ defined on (a, b) is said to be convex function if for every α such that $0 \leq \alpha \leq 1$ and for any two points x_1 and x_2 such that $a < x_1 < x_2 < b$, we have

$$f(\alpha x_1 + (1-\alpha)x_2) \leq \alpha\, f(x_1) + (1-\alpha)f(x_2) \tag{1.6.1}$$

If we put $= \frac{1}{2}$, then (1.6.1) reduces to

$$f\left(\frac{x_1 + x_2}{2}\right) \leq \frac{f(x_1) + f(x_2)}{2} \tag{1.6.2}$$

which is taken as the definition of convexity.

Remark 1.6.1: If $f''(x) \geq 0$, then $f(x)$ is convex function.

1.6.2 Strictly Convex function: A real valued function $f(x)$ defined on (a, b) is said to be strictly convex function if for every α such that $0 < \alpha < 1$ and for any two points x_1 and x_2 in (a, b), we have

$$f(\alpha x_1 + (1-\alpha)x_2) < \alpha\, f(x_1) + (1-\alpha)f(x_2) \tag{1.6.3}$$

Remark 1.6.2: If $f''(x) > 0$, then $f(x)$ is strictly convex function.

1.6.3 Concave function: A function $f(x)$ is said to be concave if $-f(x)$ is convex.

Remark 1.6.3: If $f''(x) \leq 0$, then $f(x)$ is concave function.

1.6.4 Strictly Concave function: A function $f(x)$ is said to be strictly concave if $-f(x)$ is strictly convex.

Remark 1.6.4: $f''(x) < 0$, then $f(x)$ is strictly concave function.

1.6.5 Increasing function: Let I be an open interval contained in the domain of a real function. The function $f(x)$ is an increasing function on I if $x_1 < x_2$ in I, implies

$$f(x_1) \leq f(x_2).$$

1.6.6 Decreasing function: Let I be an open interval contained in the domain of a real function. The function $f(x)$ is an decreasing function on I if $x_1 < x_2$ in I, implies

$$f(x_1) \geq f(x_2).$$

1.6.7 Maximum of a function: A function $f(x)$ is said to have a maximum value in an interval I at x_0, if $f(x_0) \geq f(x)$ for all x in I.

1.6.8 Minimum of a function: A function $f(x)$ is said to have a minimum value in an interval I at x_0, if $f(x_0) \leq f(x)$ for all x in I.

The following theorems gives the working rule for finding the points of local maxima or points of local minima. The proof is simple and hence omitted.

Theorem 1.6.1: (First derivative test) Let $f(x)$ be a differentiable function on I and let $x_0 \in I$. Then

(a) x_0 is a point of local maximum of $f(x)$ if

(i) $f'(x) = 0$

(ii) $f'(x) > 0$ at every point close to the left of x_0 and $f'(x) < 0$ at every point close to the right of x_0.

(b) x_0 is a point of local minimum of $f(x)$ if

(i) $f'(x_0) = 0$

(ii) $f'(x) < 0$ at every point close to the left of x_0 and $f'(x) > 0$ at every point close to the right of x_0.

Theorem 1.6.2: (Second derivative) Let $f(x)$ be a differential function on I and let $x_0 \in I$. Let $f''(x)$ be continuous at x_0. Then

(i) x_0 is a local maximum if both $f'(x_0) = 0$ and $f''(x_0) < 0$.

(ii) x_0 is a local minimum if both $f'(x_0) = 0$ and $f''(x_0) > 0$.

1.6.9 Gamma Function: If $n > 0$, then the integral $\int_0^\infty x^{n-1}e^{-x}dx$ which is a function of n, is called a Gamma function and is denoted by $\Gamma(n)$. Thus

$$\Gamma(n) = \int_0^\infty x^{n-1}e^{-x}dx \quad \forall\, n > 0 \tag{1.6.4}$$

1.6.10 Properties of Gamma Function: The gamma function has the following properties

(i) For $n > 1$,

$$\Gamma(n) = (n-1)\Gamma(n-1).$$

(ii) When n is a positive integer, then

$$\Gamma(n) = (n-1)!\,.$$

1.6.11 Digamma Function: The logarithmic derivative of the gamma function is called digamma function and is given by

$$\Psi(n) = \frac{d}{dn}\log\Gamma(\mathrm{n}) = \frac{\Gamma'(\mathrm{n})}{\Gamma(\mathrm{n})} \tag{1.6.5}$$

1.6.12 Beta Function: If $m, n > 0$, then the integral $\int_0^1 x^{m-1}(1-x)^{n-1}dx$, which is a function of m and n is called the beta function and is denoted by

$$\beta(m; n) = \int_0^1 x^{m-1}(1-x)^{n-1}dx \quad \forall\, m, n > 0 \qquad (1.6.6)$$

1.6.13 Properties of Beta Function: Following are the properties of Beta function

(i) Beta function is symmetric i.e., $\beta(m; n) = \beta(n; m)$.

(ii) If m, n are positive integers, then

$$\beta(m; n) = \frac{(m-1)!\,(n-1)!}{(m+n-1)!} \qquad (1.6.7)$$

Following is the relationship between Beta and Gamma functions

$$\beta(m; n) = \frac{\Gamma(m)\Gamma(n)}{\Gamma(m+n)} \qquad (1.6.8)$$

1.6.14 Leibniz Integration Rule: The Leibniz integral rule gives a formula for differentiation of a definite integral whose limits are the functions of the differential variable. It states that

$$\frac{\delta}{\delta z}\int_{a(z)}^{b(z)} f(x; z) = \int_{a(z)}^{b(z)} \frac{\delta f(x; z)}{\delta z}\,dx + f(b(z), z)\frac{\delta b(z)}{\delta z} - f(a(z), z)\frac{\delta a(z)}{\delta z} \qquad (1.6.9)$$

It is important to note that, if $a(z)$ and $b(z)$ are constants, then the last two terms of (1.6.9) vanishes.

1.7 Some Inequalities

(i) Jensen's inequality: If X is a random variable such that $E(X) = \mu$ exists and $f(x)$ is a convex function, then

$$E[f(X)] \geq f[E(X)] \qquad (1.7.1)$$

with equality iff the random variable X has a degenerate distribution at μ.

(ii) Holder's Inequality: If $x_i, y_i > 0, i = 1, 2, \ldots, n$ and $\frac{1}{p} + \frac{1}{q} = 1, p > 1$, then the following inequality holds

$$\sum_{i=1}^{n} x_i \, y_i \leq \left[\sum_{i=1}^{n} x_i^p\right]^{\frac{1}{p}} \left[\sum_{i=1}^{n} y_i^q\right]^{\frac{1}{q}} \tag{1.7.2}$$

(iii) Chebychev's Inequality: If X is a random variable with mean μ and variance σ^2, then for any positive number k

$$P\{|X - \mu| \geq k\sigma\} \leq \frac{1}{k^2} \tag{1.7.3}$$

or

$$P\{|X - \mu| < k\sigma\} \geq 1 - \frac{1}{k^2}.$$

(iv) Bienayne- Chebychev's Inequality: Let $g(x)$ be a non-negative function of a random variable X, then for any $k > 0$,

$$P\{g(x) \geq k\} \leq \frac{E[g(x)]}{k} \tag{1.7.4}$$

(v) Markov's Inequality: If we take $g(x) = |x|$ in (1.7.4), then

$$P\{|x| \geq k\} \leq \frac{E|x|}{k} \tag{1.7.5}$$

which is Markov,s Inequality.

Taking, $g(x) = |x|^r$ and replacing k by k^r in (1.7.4), we get a more generalized form of Markov's inequality

$$P\{|x|^r \geq k^r\} \leq \frac{E|x|^r}{k^r} \tag{1.7.6}$$

(vi) Log Sum Inequality: For non-negative numbers $a_1, a_2, \ldots, a_n$ and $b_1, b_2, \ldots, b_n$, the log sum inequality is given as

$$\sum_{i=1}^{n} a_i \log \frac{a_i}{b_i} \geq \sum_{i=1}^{n} a_i \log \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i} \tag{1.7.7}$$

with equality, iff $\frac{a_i}{b_i} = k$, where k is a constant.

Chapter - II

Residual Entropy and its Characterizations

Information theory is a branch of mathematics that describes how uncertainty should be quantified, manipulated and represented. Ever since the fundamental theorem of information theory were laid down by C.E. Shannon[1948], it has far reaching applications in all most every field of science and technology. Many generalizations of Shannon's entropy have been proposed by number of researchers included Arimoto [1971], Havrda and Charvat [1967], and Renyi [1961], Sharma and Mittal [1977], Sharma and Taneja [1975], Taneja [1975] and Verma [1966].

In case, one has information about the current age of the component, which can be taken into account for measuring its uncertainty, then the measure given by Shannon [1948] is not suitable. A more realistic approach which makes the use of the age into account is described by Ebrahimi [1996]. Nanda and Paul [2006] , Hooda and Kumar [2007], Abraham and Sankaran [2006] and Baig and Javid [2009,2008] introduced the concept of generalized residual entropy and apply it for some characterizations results. In this chapter, some ordering and aging properties have been defined in terms of the generalized residual entropy function and their properties have been studied.

2.1 Introduction

Recently there has been a great deal of interest in the measurement of uncertainty associated with a probability distribution. The particular interest in probability and statistics is the notion of entropy. The notion of entropy was originally developed by physicist in the context of equilibrium thermodynamics and later extended through the development of statistical mechanics. It was introduced into the information theory by C.E. Shannon [1948]. If X is a random variable having an absolutely continuous distribution function F with probability density function f, then the entropy of the random variable X is define as

$$H(X) = H(F) = -\int_{0}^{\infty} (log\, f(x))f(x)\, dx \qquad (2.1.1)$$

The entropy measures the uniformity of a distribution. As $H(f)$ increases, $f(x)$ approaches to a uniform. Consequently, the concentration of probabilities decreases and it becomes more difficult to predict an outcome of a draw from $f(x)$. In fact, a very sharply peaked distribution has a very low entropy, whereas if the probability is spread out the entropy is much higher. In this sense $H(x)$ is a measure of uncertainty associated with f.

If we think of X as the lifetime of a new unit then $H(f)$ can be useful for measuring the associated uncertainty. However, as argued by Ebrahimi $[1996]$, if a unit is known to have survived to age t, then $H(f)$ is no longer useful for measuring the uncertainty about remaining lifetime of the unit. In such situations, one should instead consider

$$H(f;t) = H(X;t) = -\int_{t}^{\infty} \left(\frac{f(x)}{\bar{F}(t)}\right)\left(log\,\frac{f(x)}{\bar{F}(x)}\right) dx$$

$$= 1 - \frac{1}{\bar{F}(t)}\int_{t}^{\infty} \big(log\lambda_F(x)\big)f(x)dx \qquad (2.1.2)$$

where $\bar{F}(x) = 1 - F(x)$ is the survival function and $\lambda_F(x) = f(x)/\bar{F}(x)$ is the hazard function of X, respectively. After the unit has survived for time t, $H(f;t)$ basically measures the expected uncertainty contained in the conditional density of $X - t$ given $X > t$ about the predictability of remaining lifetime of the unit. That is, $H(f;t)$ measures concentration of conditional probabilities.

On the basis of the measure $H(f;t)$, Ebrahimi [1996] defined and studied the following two nonparametric classes of life distributions:

Definition 2.1.1 A survival function $\bar{F}$ is said to have decreasing (increasing) uncertainty of residual life ($DURL$ ($IURL$)) if $H(f;t)$ is decreasing (increasing) in $t \geq 0$.

Ebrahimi and Kirmani [1996] explored further the properties of $DURL$ ($IURL$) classes of life distributions.

The objective of this chapter is to study the properties and implications of the dynamic measure $H(f;t)$. Preservation of $DURL$ and $IURL$ classes of life distributions under the formation of parallel systems and preservation of $DURL$ class based on record values are also discussed. We also prove and illustrate some results concerning a comparison of order statistics as well as record values on the basis of the dynamic measure $H(f;t)$. Finally, we study the characterizations of the Generalized Pareto distribution (GPD) based on $H(f;t)$. Throughout this chapter decreasing means non-increasing and increasing means non-decreasing

2.2 DURL Based on Order Statistics

An important method for increasing the reliability of a system is redundant components. A common structure of redundancy is the k-out-of-n systems. Consider a system of n components whose component lifetimes are independent and identically distributed ($i.i.d$) with common distribution F. Let the system function if and only if at least k components of n function, then the system is said to be a k-out-of-n system. Two important special cases of k-out-of-n systems can be obtained by taking $k = 1$ and $k = n$, which are known as series and parallel systems respectively.

Let $X_{1:n}, X_{2:n}, \ldots\ldots, X_{n:n}$ denote the order statistics of a set of $i.i.d.$ random variables from F. Then $X_{k:n}$ represents the lifetime of a $(n-k+1)$ out-of-n system. Also let $F_{k:n}$, $f_{k:n}$ and $\lambda_{F_{k:n}}$ denote the distribution function, the density function and the hazard rate function of $X_{k:n}$, respectively. Then,

$$f_{k:n}(x) = \frac{n!}{(k-1)!\,(n-k)!} F(x)^{k-1} \bar{F}(x)^{n-k} f(x),$$

$$F_{k:n}(x) = \sum_{i=k}^{n} \binom{n}{i} F(x)^{i} \bar{F}(x)^{n-i},$$

and

$$\lambda_{F_{k:n}}(x) = \frac{f_{k:n}(x)}{\bar{F}_{k:n}(x)}$$

$$= \frac{n!}{(k-1)!\,(n-k)!} \lambda_F(x) \frac{(F(x)/\bar{F}(x))^{k-1}}{\sum_{i=0}^{k-1}\binom{n}{i}(F(x)/\bar{F}(x))^{i}} \quad (2.2.1)$$

Lemma 2.2.1 Let X and Y be two absolutely continuous non-negative random variables with density functions f and g; hazard rates λ_{F} and λ_{G} , survival functions $\bar{F}$ and $\bar{G}$ and residual uncertainty $H(f;x)$ and $H(g;x)$, respectively. Let also θ be a non-negative increasing function such that $\lambda_{\mathrm{G}}(x) = \theta(x)\lambda_{\mathrm{F}}(x)$, $x \geq 0$ and $0 \leq \theta(x) \leq 1$.

Further, let $\lim_{x\to\infty} \bar{G}(x)/\bar{F}(x) < \infty$. If $H(f;x)$ is decreasing in x then so is $H(g;x)$.

Proof Note that from the Eq. (2.1.2), $H(g;x)$ is decreasing in x if and only if

$$E_Y[log\lambda_G(Y)|Y \geq x]$$

$$= E_Y[log\lambda_F(Y)|Y \geq x] + E_Y[log\theta(Y)|Y \geq x] \quad (2.2.2)$$

is increasing in x. Since $\theta(x)$ is increasing in x, $E_Y[log\theta(Y)|Y \geq x]$ is also increasing in x. Therefore, it is enough to show that when $m_1(x) = E_X[log\lambda_F(X)|X \geq x]$ is increasing in x then so is

$$m_2(x) = E_Y[log\lambda_F(Y)|Y \geq x].$$

Let us define the function $\beta(x)$ as follows:

$$\beta(x) = \bar{G}(x)[m_1(x) - m_2(x)] \tag{2.2.3}$$

Differentiating $\beta(x)$ with respect to x yields

$$\beta'(x) = \bar{G}(x)\{[log\lambda_F(x) - m_1(x)]\lambda_G(x) + m_1'(x)\} \tag{2.2.4}$$

Using the fact that $[m_1(x) - log\lambda_F(x)] = \left(m_1'(x)/\lambda_F(x)\right)$ we obtain

$$\beta'(x) = \bar{G}(x)m_1'(x)\left[1 - \frac{\lambda_G(x)}{\lambda_F(x)}\right] \tag{2.2.5}$$

Let $m_1(x)$ be increasing, then using the assumption that $\lambda_G(x) = \theta(x)\lambda_F(x)$ we get, $\beta'(x) > 0$. That is $\beta(x)$ is increasing in x. Now we show that m_2 is increasing. The assumption that $\lim_{x\to\infty} \bar{G}(x)/\bar{F}(x) < \infty$ implies

$$\lim_{x\to\infty} \beta(x) = \lim_{x\to\infty}\left[\frac{\bar{G}(x)}{\bar{F}(x)}\int_x^{\infty} log\lambda_F(t)f(t)dt - \int_x^{\infty} log\lambda_F(t)g(t)dt\right]$$

Hence, $\beta(x) \leq 0$ for any x, that is $m_1(x) \leq m_2(x)$. On the other hand, it is easy to show that $m_i(x),\ i = 1,2$ is increasing in x if and only if $log\lambda_F(x) \leq m_i(x),\ i = 1,2$. Hence if $log\lambda_F(x) \leq m_1(x)$ (that is $H(f;x)$ is decreasing) then $log\lambda_F(x) \leq m_2(x)$ and consequently using Eq. (2.1.2) we get that $H(g;x)$ is decreasing in x. This completes the proof.

Remark 2.2.1 Block et al. [1985] proved that when F and G are two continuous distribution functions with hazard rates λ_F and λ_G respectively, where $\lambda_G(x) = \theta(x)\lambda_F(x)$, with $0 \leq \theta(x) \leq 1$ is increasing (decreasing); then, if F is IFR (DFR), $IFRA$ $(DFRA)$, NBU (NWU) or $DMRL$ $(IMRL)$ then so is G.

The following theorem gives an important result concerning the closure property of $DURL$ distributions under formation of parallel systems.

Theorem 2.2.1 Let $X_1, X_2, \dots., X_n$ be a set of $i.i.d.$ random variables from distribution function F with density function f, hazard function λ_F and decreasing residual uncertainty $H(f;x)$. If $H(f_{n:n};x)$ denotes the residual uncertainty of the nth-order statistics of this set, then $H(f_{n:n};x)$ is also decreasing.

Proof Using Eq. (2.2.1) we have $\lambda_{F_{n:n}}(x) = \lambda_F(x)\theta(x)$, where

$$\theta(x) = n\frac{(F(x)/\bar{F}(x))^{n-1}}{\sum_{i=0}^{n-1}\binom{n}{i}(F(x)/\bar{F}(x))^i}$$

It is easy to check that $\theta(x)$ is increasing in x and has the range $(0;1)$. On the other hand, it can be easily seen that $\lim_{x\to\infty}\bar{F}_{n:n}(x)/\bar{F}(x) = 0$. Therefore, the assumptions of the Lemma 2.2.1 hold and hence $(f_{n:n}, x)$ is decreasing.

The following example shows that $IURL$ class is not preserved under formation of parallel systems.

Example 2.2.1 Consider a parallel system with two components each having unit exponential distribution. Then obviously each component is $IURL$. It can easily be shown that the failure rate of the system lifetime is given by

$$\lambda_{F_{2:2}}(x) = 1 - \frac{1}{2e^x - 1}, \quad x \geq 0,$$

which is strictly increasing Barlow and Proschan $[1981]$. Hence the system is $DURL$ and not $IURL$.

Comparison of order statistics in terms of reliability properties has also been considered in the literature. Takahashi (1988) showed that if $F_{k:n}$ is an IFR then $F_{k+1:\,n}$ is an IFR, also when $F_{k:n}$ is a DFR then so is $F_{k-1:\,n}$. Nagaraja $[1990]$ extended the Takahashi's result and showed that when $F_{k:n}$ is an IFR (DFR), an $IFRA$ $(DFRA)$ or a NBU (NWU) then the corresponding property hold for various choices of k and n. Now, using Lemma 2.2.1, we compare the order statistics in the sense of $DURL$ property.

Let us consider two sets of $i.i.d.$ random variables from the distribution function F with sizes n_1 and n_2. Let $\lambda_{F_{k_1:n_1}}$ and $\lambda_{F_{k_2:n_2}}$ denote the hazard rates $X_{k_1:n_1}$ $X_{k_2:n_2}$ respectively. Then it can be shown that $\lambda_{F_{k_2:n_2}}(x) = \theta(x)\lambda_{F_{k_1:n_1}}(x)$, such that

$$\theta(x) = \frac{c(k_1, n_1)}{c(k_2, n_2)} t^{k_2-k_1} \frac{\sum_{i=0}^{k_1-1}\binom{n}{i}t^i}{\sum_{j=1}^{k_2-1}\binom{n}{j}t^i},$$

where, $c(k,n) = n!/(n-k)!\,(k-1)!$ and $t = t(x) = F(x)/\bar{F}(x)$ is increasing in x. Nagaraja $[1990]$ proved that in the following cases $\theta(x)$ is increasing in x and its range is a subset of $(0;1)$.

- $n_1 = n_2 = n, k_1 = k, k_2 = k+1,$
- $n_1 = n, n_2 = n-1, k_1 = k_2 = k,$
- $n_1 = n, n_2 = n+1, k_1 = k, k_2 = k+1.$

Now we have the following theorem.

Theorem 2.2.2 If $X_{k:n}$ is a $DURL$; then $X_{k+1:\,n}, X_{k:\,n-1}$ and $X_{k+1:\,n+1}$ are $DURL$.

Proof The proof follows essentially from the results of Nagaraja $[1990]$ on $\theta(x)$ in these cases and Lemma 2.2.1.

In order to describe the next theorem we need the notion of record values. Let $\{X_i; i \geq 1\}$ be a sequence of $i.i.d.$ random variables with a common distribution function F which we assume to be continuous with $F(0) = 0$. The random variable X_n is called an upper record value of this sequence if $X_n > X_i$ for all $i = 1,2,\ldots,n-1$. By convention X_1 is a record value. The serial numbers at which record values occur are given by the random variables $\{L_n; n \geq 1\}$ defined recursively by $L_1 = 1$, $L_n =$

$min\{k: k > L_{n-1}, X_k > X_{Ln-1}\}, n \geq 2. \{L_n; n \geq 1\}$ is called the sequence of upper record times and $\{X_{L_n}; n \geq 1\}$ the sequence of record values correspond to $\{X_n; n \geq 1\}$. It should be noted that since $P(L_n = \infty) = 0$ for all n, the definition of $\{X_{L_n}; n \geq 1\}$ makes sense Galambos [1978].

Theorem 2.2.3 Let $\{X_n; n \geq 1\}$ be a sequence of $i.i.d.$ random variables from the distribution function F with density function f, hazard function λ_F and decreasing residual uncertainty $H(f; x)$. If f_{L_n} denote the density function of nth upper record values then $H(f_{L_n}; x)$, the residual uncertainty of f_{L_n}, is decreasing.

Proof From Arnold et al. [1992], it is known that

$$P(X_{L_n} > t) = \bar{F}(t) \sum_{k=0}^{n-1} \frac{1}{k!} [_F\ (t)]^k$$

and

$$f_{L_n}(t) = f(t) \frac{[_F(t)]^{n-1}}{(n-1)!} \qquad ,$$

where $_F\ (t) = -log\bar{F}(t)$. Combining these two equations, we get the hazard function of X_{L_n} equals to

$$\lambda_{L_n}(t) = \frac{f_{Ln(t)}}{P(X_{L_n} > t)} = \theta(x)\lambda_F(t),$$

where

$$\theta(t) = \frac{[_F\ (t)]^{n-1} / (n-1)!}{\sum_{k=0}^{n-1} \frac{1}{k!} [_F\ (t)]^k}.$$

Since $\theta(t)$ is increasing in t with range $(0,1)$, and F is $DURL$, from the Lemma 2.2.1 we get the result. This completes the proof.

Remark 2.2.2 Kochar [1990] proved that for a random variable X with a continuous distribution function F, the ratio of hazard rates of $X_{L_{n+1}}$ and X_{L_n} ,

$$\frac{\lambda_{L_{n+1}}(t)}{\lambda_{L_n}(t)} = \frac{\sum_{k=0}^{n-1} \frac{1}{k!} [_F\ (t)]^{k+1}}{n \sum_{k=0}^{n} \frac{1}{k!} [_F\ (t)]^{k}}$$

is an increasing function of t. Since this ratio has range $(0,1)$, using Lemma 2.2.1 we conclude that when X_{L_n} is $DURL$ then so is $X_{L_{n+1}}$.

2.3 Characterization Results

2.3.1 Characterizations from the residual entropy

The general characterization problem is to determine when the residual entropy uniquely determines the distribution function. Ebrahimi [1996] and Rajesh and Nair [1998] showed that $H(t)$ uniquely determines the distribution function $F(t)$ in absolutely continuous and discrete cases, respectively.

$(\boldsymbol{a})$ Continuous Case

Theorem 2.3.1 If X has an absolutely continuous distribution $F(t)$ and an increasing residual entropy $H(t)$, then $H(t)$ uniquely determines $F(t)$.

Proof From the definition, we have

$$H(t) = -\int_t^\infty \frac{f(x)}{1-F(t)} log \frac{f(x)}{1-F(t)} dx$$

which is equivalent to

$$\int_t^\infty f(x) log f(x) dx = \left(1-F(t)\right) log \left(1-F(t)\right) - \left(1-F(t)\right) H(t)$$

Differentiating, we obtain

$$f(t) log f(t) = f(t) \left(1 - H(t) + log\left(1-F(t)\right)\right) + \left(1-F(t)\right) H'(t).$$

Thus, the failure rate $r(t) = f(t)/(1-F(t))$ verifies

$$r(t)(H(t) - 1 + logr(t)) = H'(t) \tag{2.3.1}$$

Hence, for a fixed $t > 0$, $r(t)$ is a positive solution of the following equation

$$g(x) = x(H(t) - 1 + logx) - H'(t) = 0 \tag{2.3.2}$$

Note that $g(0) = -H'(t) \leq 0, g(+\infty) = +\infty$ and

$$g'(X) = H(t) + logx,$$

so, $g(x)$ first decreases and then increases in x, with a minimum at $x_t = e^{-H(t)}$, which implies that equation (2.3.2) has a unique positive solution $(r(t))$ for all t.Thus, H uniquely determines r and hence, F .

Remark 2.3.1 Note that for every $t \geq 0$, equation (2.3.2) has also a unique solution if $g(x_t) = 0$, which gives $H(t) = log(b - t)$, i.e. the residual entropy for the Uniform distribution in $(0, b)$. Thus, the Uniform distribution, can be characterized from a decreasing residual entropy $H(t) = log\ (b - t)$.

Remark 2.3.2 If $H(t)$ is decreasing and $g(x_t) \neq 0$, then (2.3.2) has two solutions, for all $t \geq 0$. One of them is the failure rate $r(t)$ and the other one could be not a proper failure rate. Next, we see an example where both solutions are failure rates.

Example 2.3.1 If X has a Beta $\beta(c, 1)$ (Power) distribution $F(t) = t^c$ for $0 < t < 1$ and $c > 1$, then $r(t) = ct^{c-1}/(1 - t^c)$ and

$$H(t) = \frac{c-1}{c} + lo\ g\left(\frac{1-t^c}{c}\right) + \frac{(c-1)t^c}{1-t^c} logt$$

Thus, g has a minimum at

$$x_t = \frac{c}{1-t^c} t^{-(c-1)t^c/(1-t^c)} exp\left\{-\frac{c-1}{c}\right\},$$

which verifies

$$\frac{x_t}{r(t)} = t^{-(c-1)/(1-t^c)} exp\left\{-\frac{c-1}{c}\right\} > 1 \quad \text{for } 0 < t < 1$$

So, for every $t \geq 0$, equation (2.3.2) has two positive solutions $r_1(t) < x_t < r_2(t)$, where $r_1(t)$ is the failure rate of the Beta distribution. Hence, the other solution $r_2(t)$, is also a failure rate function $(i.e. r_2 \geq 0 \; and \; \int_0^1 r_2 = \infty)$.

(b) Discrete Case

Theorem 2.3.2 If X has a discrete distribution $F(t)$ with support $\{t_j : t_j < t_{j+1}\}$ and an increasing residual entropy $H(t)$, then $H(t)$ uniquely determines $F(t)$.

Proof From the definition of $H(t)$ in the discrete case, we have

$$H(t_j) = -\sum_{k=j}^{\infty} \frac{p(t_k)}{\bar{F}(t_j)} log \frac{p(t_k)}{\bar{F}(t_j)}$$

which is equivalent to

$$\sum_{k=j}^{\infty} \mathrm{p(t_k)} log \mathrm{p(t_k)} = \bar{\mathrm{F}}(\mathrm{t_j}) log\left(\bar{\mathrm{F}}(\mathrm{t_j})\right) - \bar{\mathrm{F}}(\mathrm{t_j})\mathrm{H}(\mathrm{t_j}).$$

For t_{j+1} we obtain

$$\sum_{k=j+1}^{\infty} \mathrm{p(t_k)} log \mathrm{p(t_k)} = \bar{\mathrm{F}}(\mathrm{t_{j+1}}) log\left(\bar{\mathrm{F}}(\mathrm{t_{j+1}})\right) - \bar{\mathrm{F}}(\mathrm{t_{j+1}})\mathrm{H}(\mathrm{t_{j+1}})$$

By subtracting, we obtain

$$p(t_j) logp(t_j) = \bar{F}(t_j)\left(log\left(\bar{F}(t_j)\right) - H(t_j)\right) - \bar{F}(t_{j+1})\left(log\left(\bar{F}(t_{j+1})\right) - H(t_{j+1})\right),$$

which, using $p(t_j) = \bar{F}(t_j) - \bar{F}(t_{j+1})$, is equivalent to

$$\bar{F}(t_j)\left(logp(t_j) - log\left(\bar{F}(t_j)\right) + H(t_j)\right)$$
$$= \bar{F}(t_{j+1})(logp(t_j) - log\left(\bar{F}(t_{j+1})\right) + H(t_{j+1}))$$

If $\lambda_j = \bar{F}(t_{j+1})/\bar{F}(t_j)$, then

$$H(t_j) + log(1 - \lambda_j) = \lambda_j \; (H(t_{j+1}) + log \; (\frac{1}{\lambda_j} - 1))$$

holds. Hence, λ_j is a number in $(0,1)$ which is a solution of the following equation

$$g(x) = \; -xH(t_{j+1}) + H(t_j) + xlog(x) + (1 - x) \; log(1 - x)$$
$$= 0 \qquad (2.3.3)$$

In this case $g(0) = H(t_j) \geq 0, g(1) = H(t_j) - H(t_{j+1}) \leq 0$ and

$$g'(x) = \; -H(t_{j+1}) + log \left(\frac{x}{1 - x}\right).$$

Therefore, $g(x)$ first decreases and then increases in $(0,1)$ with a minimum at $x_j = 1/(1 + e^{-H(t_{j+1})})$, which implies that equation (2.3.3) has a unique positive solution (λ_j) in $(0,1)$ for all j. Finally, since $1 - \lambda_j = r(t_j)$ where $r(t_j) = p(t_j)/\bar{F}(t_j)$ is the discrete failure rate, then H uniquely determines r (or F).

Remark 2.3.3 Note that (2.3.3) has also a unique solution if $g(x_j) = 0$, which gives $H(t) = log(n - t + 1)$, for $t = 1,2, \dots, n$. i.e. the residual entropy for the discrete Uniform distribution in $1,2, \dots, n$. Thus, the Uniform distribution, can be characterized from a decreasing residual entropy.

Remark 2.3.4 In general, $H(t)$ does not uniquely determine $F(t)$ For example, if X has a Bernoulli distribution $B(p)$, then

$$p(t) = \; Pr(X = t) = \begin{cases} 1 - p & if \; t = 0 \\ p & if \; t = 1 \end{cases}$$

and

$$\mathrm{H(t)} = \begin{cases} -(1 - p) \; log(1 - p) - p log p & if \; t = o \\ 0 & if \; t = 1 \end{cases}$$

Note that $H(t)$ is decreasing. Thus, the residual entropy $H(t)$ is the same for Bernoulli's distributions $B(p)$ and $B(1-p)$.

Remark 2.3.5 In particular, if $H(t)$ is constant, then from theorems 2.3.1 and 2.3.2, we obtain the characterizations given by Ord et al. [1983] for the Exponential and Geometric distributions in continuous and discrete cases, respectively. In Tables 1 and 2, the residual entropy for some distributions are given. Note that, from Tables 1 and 2, we can characterize the Exponential, Weibull (with shape Parameter $b > 1$), Pareto and Geometric distributions because $H(t)$ is increasing. We can also characterize the continuous and discrete Uniform distributions from the preceding remarks. However, we can not characterize the Beta or the Finite Range distributions.

Remark 2.3.6 Replacing X by $-X$, the analogous results to the reversed residual entropy, that is the residual entropy for right truncated distributions, $H(X|X \leq t)$, can be obtained

Table 1. Residual entropy: continuous case

Model	$\boldsymbol{f(t)}$	$\boldsymbol{t}$	$\boldsymbol{H(t)}$

Uniform	$\frac{1}{b-a}$	$a < t < b$	$log(b-t)$
Exponential	$aexp(-at)$	$t > 0, a > 0$	$1 - loga$
Weibull	$ab(at)^{b-1}exp(-(at)^b)$	$t > 0$ $a, b \; 0$	$\nearrow b \; 1$ $\searrow b \; 1$
Pareto	$\frac{1}{a+bt}\left(\frac{a}{a+bt}\right)^{\frac{1}{b}}$	$t > 0, a, b > 0$	$b + 1$ $+ log(a+bt)$
Finite Range	$\frac{1}{a+bt}\left(\frac{a}{a+bt}\right)^{\frac{1}{b}}$	$0 < t < -a/b$ $a > 0, b < 0$	$b + 1$ $+ log(a+bt)$
Beta$(c, 1)$	ct^{c-1}	$0 < t < 1,$ $c > 0$	$\frac{c-1}{c} + log\left(\frac{1-t^c}{c}\right)$ $+ \frac{(c-1)t^c}{1-t^c}logt$

Table 2. Residual entropy: discrete case

Model	$\boldsymbol{p(t)}$	$\boldsymbol{t}$	$\boldsymbol{H(X)}$	$\boldsymbol{H(t)}$
Uniform	$\frac{1}{n}$	$1, \dots, n$	$logn$	$log(n-t+1)$
Geometric	pq^t	$0,1, \dots$	$-logp - \frac{q}{p}logq$	$-logp - \frac{q}{p}logq$

2.3.2 Characterization of the generalized pareto distribution based on residual uncertainty

In reliability theory, in studies of the lifetime of a component or a system, a flexible model which has been widely used in the literature is that of a Generalized Pareto distribution (GPD) with survival function,

$$\bar{F}(x) = \left(\frac{b}{ax+b}\right)^{1/a+1}, \quad x \geq 0 \qquad (2.3.4)$$

where $a > -1$ and $b > 0$. This model has been considered, among others, by Hall and Wellner $[1981]$, and includes the exponential distribution $(a \to 0)$, Lomax (Pareto) distribution $(0 < a)$ and the power distribution $(-1 < a < 0)$. In the following theorems, we present some results characterizing the GPD based on residual uncertainty.

Theorem 2.3.3 Let X be a non-negative absolutely continuous random variable with survival function $\bar{F}(x)$, hazard rate $\lambda_F(x)$ and residual uncertainty $H(f;x)$, where f is density function of X. Then,

$$H(f;x) = c - log\lambda_F(x) \qquad (2.3.5)$$

if and only if F is GPD with survival function of form $(2.3.4)$; where c is a real-valued constant.

Proof The if' part of the theorem is straightforward. To prove the `only if ' part let $(2.3.5)$ be valid. This is equivalent to

$$(c-1)\bar{F}(x) + \int_x^\infty f(t)log\lambda_F(t)dt = \bar{F}(x)log\lambda_F(x) \qquad (2.3.6)$$

Differentiating both sides of Eq.$(2.3.6)$ with respect to x implies

$$\frac{\lambda'_F(x)}{\lambda_F^2(x)} = -(c-1).$$

Solving this differential equation yields

$$\lambda_F(x) = \frac{1}{(c-1)x+d},$$

where $d^{-1} = \lambda_F(0)$. This is the hazard function of the GPD. Since the distribution function is uniquely determined by the hazard rate, the proof is complete.

It is generally known that the mean residual life function $\delta_F(t), \delta_F(t) = E(X-t|X > t)$, is not the same as $1/\lambda_F(t)$. The following theorem gives another characterization of GPD.

Theorem 2.3.4 Let X be a non-negative absolutely continuous random variable with survival function $\bar{F}(x)$, the mean residual life function $\delta_F(x)$, and residual uncertainty $H(f;x)$. Then,

$$H(f;x) = c + log\delta_F(x) \tag{2.3.7}$$

if and only if F is GPD of the form (2.3.4), where c is a real valued constant.

Proof The if' part of the theorem is easy to prove. To prove the only if part note that in general one can easily verify that the derivative of $H(f;x)$ with respect to x is

$$H'(f;x) = \lambda_F(x)[H(f;x) + log\lambda_F(x) - 1] \tag{2.3.8}$$

From (2.3.7) and (2.3.8) we have

$$\frac{\delta_F'(x)}{\delta_F(x)} = \lambda_F(x)[H(f;x) + log\lambda_F(x) - 1]$$

$$= \lambda_F(x)[c + log\delta_F(x)\lambda_F(x) - 1] \tag{2.3.9}$$

Using the fact that $\delta_F'(x) = -1 + \lambda_F(x)\delta_F(x)$, we can easily see that $\delta_F(x)$ is a linear function. That is F is GPD.

2.3.3 Characterizations from the residual gamma-entropy

The preceding results can be extended to a more general information measure, the γ −entropy, $\gamma > -1$ (Rao [1965] and Ord et al. [1981]) defined by

$$H_\gamma(X) = \frac{1}{\gamma}\int f(t)(1 - f^\gamma(t))dt \tag{2.3.10}$$

The Shannon's entropy can be obtained as $H(X) = \lim_{\gamma\to 0} H_\gamma(X)$.

Analogously, we can define the residual (left truncated) γ −entropy by

$$\begin{aligned} H_\gamma(t) &= H_\gamma(X - t|X \geq t) \\ &= \frac{1}{\gamma}\int_t^\infty \frac{f(x)}{\bar{F}(t)}\left(1 - \frac{f^\gamma(x)}{\bar{F}^\gamma(t)}\right)dx \end{aligned} \tag{2.3.11}$$

for $t \in D = \{x \in \mathbb{R}|\bar{F}(x) > 0\}$, and the right or the interval truncated γ −entropy. For this measure we have the following results.

(a) Continuous Case

Theorem 2.3.5 If X has an absolutely continuous distribution $F(t)$ and an increasing $H_\gamma(t)$ then $H_\gamma(t)$ uniquely determines $F(t)$.

Proof From the definition, we have

$$\begin{aligned} n(t) &= 1 - \gamma H_\gamma(t) \\ &= \frac{1}{\bar{F}^{\gamma+1}(t)}\int_t^\infty f^{\gamma+1}(x)dx > 0 \end{aligned}$$

which is equivalent to

$$\bar{F}^{\gamma+1}(t)n(t) = \int_t^\infty f^{\gamma+1}(x)dx$$

and differentiating, we obtain

$$-(\gamma + 1)\bar{F}^\gamma(t)f(t)n(t) + \bar{F}^{\gamma+1}(t)n'(t) = -f^{\gamma+1}(t).$$

Thus, the failure rate $r(t) = f(t)/\bar{F}(t)$ verifies

$$-(\gamma + 1)r(t)n(t) + n'(t) = -r^{\gamma+1}(t).$$

Hence, $r(t)$ is a positive solution of the following equation

$$g'(x) = x^{\gamma+1} - (\gamma+1)n(t)x + n'(t) = 0 \qquad (2.3.12)$$

where

$g'(x) = (\gamma+1)x^{\gamma} - (\gamma+1)n(t)$.

Moreover, if $\gamma > 0(\gamma < 0)$, then $g(0) = n'(t) \leq 0(\geq)$,

$g(+\infty) = +\infty(-\infty)$. So, $g(x)$ first decreases (increases) and then increases (decreases) in x, with a minimum (maximum) at $x_t = n^{1/\gamma}(t)$, which implies that equation (2.3.12) has a unique positive solution $(r(t))$ for all t.

Remark 2.3.7 Note that equation (2.3.12) has also a unique solution if $g(x_t) = 0$, for all $t \geq 0$, which gives $H_\gamma(t) = (1-(b-t)^{-\gamma})/\gamma$, i.e. the residual γ −entropy for the Uniform distribution. Thus, the Uniform distribution, can be characterized from a decreasing $\gamma -$ residual entropy.

(*b*) Discrete Case

Theorem 2.3.6 If X has a discrete distribution $F(t)$ with support $\{t_j : t_j < t_{j+1}\}$ and $H_\gamma(t)$ is increasing in t, then $H_\gamma(t)$ uniquely determines $F(t)$.

Proof From the definition of $H_\gamma(t)$, we have

$$n(t_j) = 1 - \gamma H_\gamma(t_j) = \frac{1}{\bar{F}^{\gamma+1}(t_j)} \sum_{k=j}^{\infty} p^{\gamma+1}(t_k) > 0$$

which implies

$$p^{\gamma+1}(t_j) = \bar{F}^{\gamma+1}(t_j)n(t_j) - \bar{F}^{\gamma+1}(t_{j+1})n(t_{j+1}),$$

which, using $p(t_j) = \bar{F}(t_j) - \bar{F}(t_{j+1})$, is equivalent to

$$\left(\bar{F}(t_j) - \bar{F}(t_{j+1})\right)^{\gamma+1} = \bar{F}^{\gamma+1}(t_j)n(t_j) - \bar{F}^{\gamma+1}(t_{j+1})n(t_{j+1}).$$

If $\lambda_{j\,=}\ \bar{F}(t_{j+1})/\bar{F}(t_j)$ then

$$\left(1-\lambda_j\right)^{\gamma+1} = n\left(t_j\right) - \lambda_j^{\gamma+1} n\left(t_{j+1}\right)$$

holds. Hence, λ_j is a number in $(0,1)$ which is a solution of the following equation

$$g(x) = x^{\gamma+1} n\left(t_{j+1}\right) + (1-x)^{\gamma+1} - n\left(t_j\right) = 0 \qquad (2.3.13)$$

where if $\gamma > 0$ $(\gamma < 0)$ and $H_\gamma(t)$ is increasing, then $g(0) = 1 - n\left(t_j\right) \geq 0(\leq)$, $g(1) = n\left(t_{j+1}\right) - n\left(t_j\right) \leq 0(\geq)$ and

$$g'(x) = (\gamma+1)x^\gamma n\left(t_{j+1}\right) - (\gamma+1)(1-x)^\gamma.$$

So, $g(x)$ first decreases (increases) and then increases (decreases) in $(0,1)$, with a minimum at $x_j = 1/\left(1 + n^{1/\gamma}\left(t_{j+1}\right)\right)$, which implies that equation (2.3.13) has a unique positive solution $\left(\lambda_j\right)$ in $(0,1)$ for every t_j. Finally, since $1 - \lambda_j = r(t_j)$ where, for every t_j, $r\left(t_j\right) = p(t_j)/\bar{F}\left(t_j\right)$ is the discrete failure rate, H_γ uniquely determines r $(or\ F\ )$.

Remark 2.3.8 Again, note that equation (2.3.13) has a unique solution if $g\left(x_j\right) = 0$, and characterizes the discrete Uniform distribution from a decreasing discrete $\gamma -$ residual distribution.

Remark 2.3.9 In general, $H_\gamma(t)$ does not determine $F(t)$. For example, in the Bernoulli distribution case. Another example is the case $\gamma = 1$ and $F(t) = 1 - (1-t)^a$, and for $0 < t < 1$ and $a = 2 \pm \sqrt{2}$.

Remark 2.3.10 In the special case of $\gamma = 1$, we can give an inversion formula to obtain F from an increasing function H_1, by using the following equalities

$$r(t) = 1 - H_1(t) + \sqrt{\left(1 - H_1(t)\right)^2 + H_1'(t)}$$

$$F(t) = 1 - exp\left\{-\int_{-\infty}^{t} r(x)dx\right\}$$

(in the continuous case). A similar formula can be obtained when $\gamma = 2$. We can also obtain the inversion formulas for H_1 and H_2 in the discrete case.

Remark 2.3.11 Replacing X by $-X$, the analogous results for the reversed residual $\gamma -$ entropy can be obtained.

2.4 Characterizations by Relationships

Asadi and Ebrahimi [2000] characterized the Generalized Pareto Distribution (GPD) from the equality $H(t) = c - \log r(t)$ where c is a real-valued constant. This characterization includes the Exponential distribution $(c = 1)$, the Pareto $(c > 1)$ and the Finite Range $(c < 1)$. Belzunce et al. [2004] gave the more general case where c is a function of t.

Theorem 2.4.1 If X has an absolutely continuous distribution $F(t)$, with support (α, β), and $H(t) = c(t) - logr(t)$, then

$$r(t) = \frac{1}{K - \int_{\alpha}^{t} e^{c(x)}(1 - c(x))dx} e^{c(t)} \tag{2.4.1}$$

where $K = e^{H(\alpha)}$.

Proof From residual entropy's definition, we have

$$H(t) = 1 - \frac{1}{1 - F(t)} \int_{t}^{\infty} f(x) logr(x) dx \,,$$

which jointly with $H(t) = c(t) - logr(t)$, gives

$$r'(t) = c'(t)r(t) + \big(1 - c(t)\big)r^2(t).$$

Solving this Bernoulli's differential equation we obtain

$$\frac{1}{r(t)} = e^{-c(t)} \left(K - \int_{\alpha}^{t} e^{c(x)}(1 - c(x))dx\right),$$

and hence (2.4.1).

Remark 2.4.1 In particular, if $c(t)$ is constant, then we obtain the characterization given by Asadi and Ebrahimi [2000] for the GPD which includes Exponential, Pareto and Finite Range distributions.

Remark 2.4.2 From Theorem 2.4.1, we can characterize new distribution models. For example, if $c(t) = at + b$ for $t > 0$ and $a, b > 0$, then we obtain the general model with failure rate

$$r(t) = \frac{a}{\left(2 - b + \frac{a}{b}\right)e^{-at} + at + b - 2}.$$

We have to check if $r(t)$ is a failure rate $\left(i.e\ r \geq 0\ and\ \int_0^\infty r = \infty\right)$. It is not difficult to see that if $(2 - b + a/b) \leq 0\ or\ (2 - b + a/b) > e^{1-b}$, then r is a failure rate.

Nair and Rajesh [1998] characterized the Exponential distribution from the equality $H(t) + e(t) = H(0) + e(0)$ (where $e(t)$ is the mean residual life function). They also characterized the type ***I*** Extreme Value distribution (defined by $F(t) = 1 - \exp(-\exp(t))$, for all $t \in \mathbb{R}$) from $H(t) + e(t) = 1 - t$.

Theorem 2.4.2 If X has an absolutely continuous distribution $F(t)$, $H(t) + e(t) = c(t)$ and $c'(t) \geq -1$, then $c(t)$ uniquely determines $F(t)$.

Proof Differentiating $H(t) + e(t) = c(t)$, we have

$$H'(t) + e'(t) = c'(t)$$

which jointly with $r(t) = (e'(t) + 1)/e(t)$, gives

$$H'(t) + e(t)r(t) - 1 = c'(t)$$

and, from (2.3.1),

$$r(t)H(t) - r(t) + r(t)logr(t) + e(t)r(t) = c'(t) + 1$$

holds. By using again $H(t) + e(t) = c(t)$, we obtain

$$r(t)\left(c(t) - 1 + logr(t)\right) = c'(t) + 1$$

Thus, to obtain $r(t)$ we must solve the equation

$$g(x) = c'(t) + 1 \tag{2.4.2}$$

where

$$g(x) = x(c(t) - 1 + logx)$$

This equation is similar to $(2.3.2)$ and it is easy to show that $g(0) = 0$ and that $g(x)$ first decreases and then increases in x. Hence, equation $(2.4.2)$ has a unique solution (for each t) when $c'(t) \geq -1$ and $c(t)$ uniquely determines $r(t)$.

Remark 2.4.3 In particular, we obtain the characterizations given by Nair and Rajesh [1998] for the Exponential and Extreme Value distributions when $c(t)$ is constant or $c(t) = 1 - t$.

Chapter - III

Past Residual Entropy and its Generalization

3.1 Introduction

Let X be a non-negative random variable having the cumulative distribution function $F(x)$, and the survival function $\bar{F}(x) = 1 - F(x)$. The basic measure of uncertainty of a random variable X with density function $f(x)$ is given by Shannon [1948] and is defined as

$$H(X) = -\int_{0}^{\infty} f(x) log f(x) dx. \qquad (3.1.1)$$

In recent years, the role of Shannon entropy as a measure of uncertainty in residual lifetime distributions has been studied by many researchers including Ebrahimi [1996], Ebrahimi and Kirmani [1996], Ebrahimi and Pellerey [1995], Belzunce et al. [2004].

Ebrahimi [1996] defined the uncertainty of the residual life time distributions, $H(X, t)$ of a component t as

$$H(X,t) = -\int_{t}^{\infty} \frac{f(x)}{\bar{F}(t)} log \frac{f(x)}{\bar{F}(t)} dx$$

$$= log \bar{F}(t) - \frac{1}{\bar{F}(t)} \int_{t}^{\infty} f(x) log f(x) dx$$

$$= 1 - \frac{1}{\bar{F}(t)} \int_{t}^{\infty} f(x) log h(x) dx \qquad (3.1.2)$$

where $h(t) = f(t)/\bar{F}(t)$ is the hazard rate or failure rate function and $\bar{F}(t)$be the reliability function of the random variable X.

It is reasonable to presume that in many realistic situations uncertainty is not necessary related to the future but can also refer to the past. For instance, if at time, a system which is observed only at certain pre-assigned inspection times, is found to be down, then the uncertainty of the system life relies on the past, i.e., on which instant in $(0, t)$ it has failed. Based on this idea, Crescenzo and Longobardi [2002] have studied the past entropy over $(0, t)$. If X denotes the life time of a component or of living organism, then the past entropy of X is defined as

$$\bar{H}(X,t) = -\int_0^t \frac{f(x)}{F(t)} log \frac{f(x)}{F(t)} dx$$

$$= 1 - \frac{1}{F(t)} \int_0^t f(x) log\tau(x) dx \qquad (3.1.3)$$

where $F(t)$ is the cumulative distribution function and $\tau(x) = \frac{f(x)}{F(t)}$ is the reversed hazard function or reserved failure rate of X. The function $\tau(x)$ is receiving increasing attention in reliability theory and survival analysis. For more properties and applications of reversed hazard rate function, one should refer to Chandra and Roy [2001] and Sankaran and Gleeja [2006].

The following example shows the role of the past entropy in the comparison of random lifetimes.

Example 3.1 Let X and Y be two non-negative random variables during the lifetimes of the two components with p.d.f's

$$f_X(t) = 2t, \qquad 0 < t < 1$$

and

$$f_Y(t) = 2(1-t), \quad 0 < t < 1$$

respectively.

The differential entropy of the random variable X is given by

$$H(X) = -\int_0^1 2xlog2xdx$$

$$= log2 - \frac{1}{2}$$

and that of Y is given by

$$H(Y) = log2 - \frac{1}{2}.$$

Therefore, the differential entropy of the two random variable is equal, so that the expected uncertainty contained in f_X and f_Y about the predictability of the outcomes of X and Y is the same. However, if both components are found to have failed upon an inspection at time $t \epsilon (0,1)$, then the uncertainty about their unknown failure times must be measured by means of the past entropy. The past entropy of the random variables X and Y is obtained by using (3.1.3).

Therefore,

$$\overline{H}(X,t) = \frac{1}{2} + log\frac{t}{2},$$

and

$$\overline{H}(Y,t) = \frac{1}{2} + \frac{(1-t)^2}{1-(1-t)^2}log\frac{2(1-t)}{1-(1-t)^2} - \frac{1}{1-(1-t)^2}log\frac{2}{1-(1-t)^2},$$

so that $\overline{H}(X,t) \geq \overline{H}(Y,t)$ for all $t \epsilon (0,1)$. Hence, even though $H(X) = H(Y)$, the expected uncertainty contained in the p.d.f of X given $X \leq t$ about the predictability of the failure time of the first component is larger than the expected uncertainty contained in the p.d.f of Y given $Y \leq t$ about the predictability of the failure time of the second component.

Given that at time t a component is found to be down $\overline{H}(X,t)$ measures the uncertainty about its past life. In forensic science a lifetime distribution truncated above time t is of utmost importance.

In the following section we express the entropy of a random lifetime in terms of the residual entropy and of the past entropy. The problem of the increasing nature of $\bar{H}(t)$ is also discussed. In particular, we show that a random lifetime X has increasing past entropy if its reversed failure rate is decreasing, $i.e. X$ is $DRFR$. Upper bounds for $\bar{H}(t)$ and for the reversed hazard function are also obtained. Finally, we analyse the effect of strict monotonic transformations on the past entropy, and particularly their effects on the entropy monotonicity properties.

Throughout this chapter, the terms decreasing and increasing are used in a non-strict sense.

3.2 Results on the Past Entropy

From (3.1.3) we also have the following expressions for the past entropy:

$$\bar{H}(t) = logF(t) - \frac{1}{F(t)}\int_0^t f(x)logf(x)dx$$

$$= 1 - \frac{1}{F(t)}\int_0^t f(x)log\tau(x)dx \qquad (3.2.1)$$

where $\tau(t) = f(t)/F(t)$ is the reversed hazard function, or reversed failure rate, of X. The function $\tau(t)$ is receiving increasing attention in reliability theory and survival analysis (Block et al. [1998] and Chandra and Roy [2001]). As pointed out by some authors (in particular, Nanda and Shaked [2001]), its role is dual to that of $r(t)$. Indeed, as will appear clear in the following, the role of $\tau(t)$ in the analysis of the past entropy is analogous to that of $r(t)$ in the analysis of the residual entropy as performed by Ebrahimi [1996].

Throughout the chapter we make use of the following relation, which is an immediate consequence of (3.2.1):

$$\frac{d}{dt}\bar{H}(t) = \tau(t)[1 - \bar{H}(t) - log\tau(t)]. \qquad (3.2.2)$$

In the following proposition we show that $(3.1.1)$ can be expressed in terms of $H(t)$ and $\bar{H}(t)$. The proof is omitted, being straightforward.

Proposition 3.2.1 For all $t > 0$,

$$H(X) = H[F(t), \bar{F}(t)] + F(t)\bar{H}(X,t) + \bar{F}(t)H(X,t) \quad (3.2.3)$$

where $\mathrm{H}[p, 1-p] = -plogp - (1-p)log(1-p)$ is the entropy of a Bernoulli distribution.

Proof The proof follows by substituting the value of $H[F(t), \bar{F}(t)]$, $\bar{H}(X,t)$ and $H(X,t)$ on the R.H.S of $(3.2.3)$.

The identity $(3.2.3)$ has the following interpretation.

The uncertainty about the failure time of an item can be decomposed into three parts:

(i) the uncertainty of whether the item has failed before or after time t,

(ii) the uncertainty about the failure time in $(0, t)$ given that the item has failed before t, and

(iii) the uncertainty about the failure time in $(t, +\infty)$ given that the item has failed after t.

From $(3.1.3)$ it is immediately seen that the entropy of a random variable uniformly distributed on $(0, t)$ is given by $logt$. The latter is also an upper bound for $\bar{H}(t)$; indeed

$$\bar{H}(t) \leq logt \quad \text{for all } t > 0 \quad (3.2.4)$$

This is in agreement with the principle of maximum entropy $($Ebrahami $[2000])$, according to which the uniform distribution, maximizes entropy under the constraint that the probability mass is concentrated on a finite interval. A direct consequence of $(3.2.4)$ is that the past entropy of a random lifetime cannot be constant for all $t > 0$.

Similarly to (3.2.4) we can prove that the past entropy of a random lifetime X distributed over $(0,b)$ satisfies $\overline{H}_X(t) \leq \overline{H}_{U(0,b)}(t)$, where

$$\overline{H}_{U(0,b)}(t) = \begin{cases} logt & if\ 0 < t \leq b, \\ logb & if\ t > b, \end{cases} \tag{3.2.5}$$

is the past entropy of a random variable uniformly distributed over $(0,b)$.

Note that, $logt$ is not always a tight bound for the past entropy, especially for large t. If, for instance, X is exponentially distributed with mean $1/\lambda$, then we have

$$\overline{H}(t) = 1 + log\left(1 - e^{-\lambda t}\right) - \frac{\lambda t e^{-\lambda t}}{1 - e^{-\lambda t}}, \qquad t > 0.$$

while, if X has a Pareto-type CDF $F(t) = (t/(1+t))1_{(0,+\infty)}(t)$, it is given by

$$\overline{H}(t) = 1 + log\frac{t}{1+t} - \frac{1}{t}log(1+t)^2, \qquad t > 0.$$

In both cases $\lim_{t\to\infty} \overline{H}(t) = 1$, while $logt$ diverges as t goes to $+\infty$.

Let us now discuss a problem concerning a conditioned mean value of a random lifetime X. By setting $\mu(t) = E(X \mid X \leq t)$ it is not hard to see that $(d/dt)\mu(t) = \tau(t)[t - \mu(t)] \geq 0$, so that the mean failure time conditioned by a failure before t is increasing in $t \geq 0$. We emphasize that a relevant difference exists between the past entropy and the conditioned mean value: while $\mu(t)$ is always increasing, the past entropy is not necessarily increasing, as shown in the following example in which neither $\overline{H}(t)$ nor $\tau(t)$ is monotonic in t.

Example 3.2.1 Let X be a random lifetime having c.d.f

$$F(t) = exp\left\{-1 - \frac{1}{t}\right\}\mathbf{1}_{(0,1]}(t) + exp\left\{\frac{t^2 - 5}{2}\right\}\mathbf{1}_{(1,2]}(t)$$
$$+ exp\left\{-\frac{1}{t}\right\}\mathbf{1}_{2,\infty}(t). \tag{3.2.6}$$

Then

$$\tau(t) = \frac{1}{t^2}\mathbf{1}_{(0,1]\cup(2,+\infty)}(t) + t\mathbf{1}_{(1,2]}(t),$$

so that $\tau(t)$ is not monotonic. Moreover, we can also prove that the past entropy is not monotonic, as shown in Figure 1.

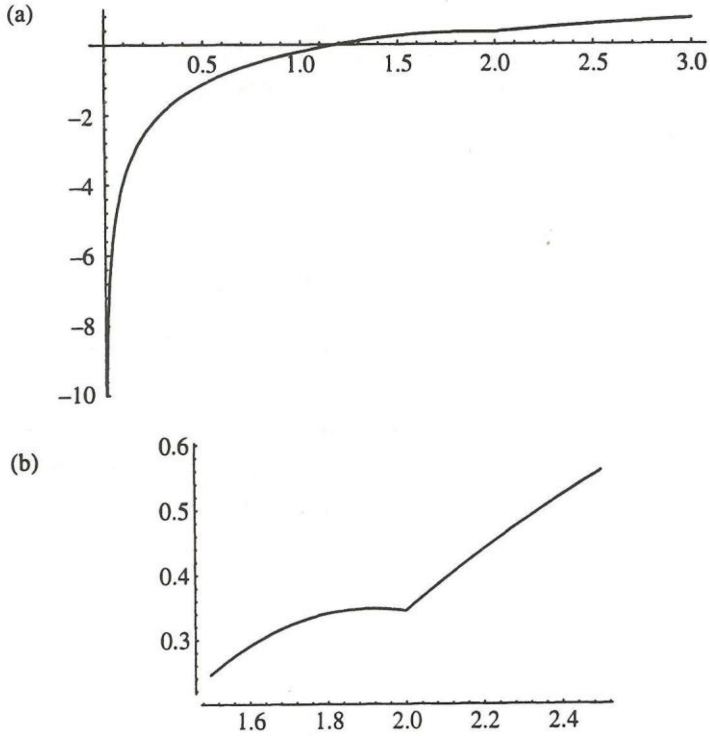

Figure 1: (a) The past entropy corresponding to (3.2.6) is sketched for t∈ (0,3).
(b) An enlargement of (a) close to t=2, making it evident that it is not monotonic.

It is not hard to prove that if $f(t)$ is decreasing in $t > 0$, then $\bar{H}(t)$ is increasing in $t > 0$. However, this property can be proved under the weaker assumption that the reversed failure rate of X is decreasing $(\text{i.e.}\, X \text{ is } DFRF)$. Indeed, if $\tau(t)$ is decreasing, from $(3.2.1)$ we have $\bar{H}(t) \le 1 - log\tau(t)$, which gives the following result.

Proposition 3.2.2 If $\tau(t)$ is decreasing for all $t > 0$, then $\bar{H}(t)$ is increasing for all $t > 0$.

Remark 3.2.1 If X is a random lifetime with increasing failure rate and decreasing reversed failure rate, then the second term of the right-hand side of (3.2.3) is increasing in $t > 0$ (due to Proposition 3.2.2), while the third term is decreasing in $t > 0$ (Ebrahimi [1997]).

The two upper bounds in the following proposition can be obtained by making use of (3.2.2).

Proposition 3.2.3 If $\bar{H}(t)$ is increasing for $t > 0$, then $\tau(t) \leq exp\{1 - \bar{H}(t)\}, t > 0$ and

$$\bar{H}(t) \leq 1 - log\, \tau(t), t > 0 \qquad (3.2.7)$$

Remarks 3.2.2

(i) From (3.2.2) and (3.2.4) we have that $\bar{H}(t)$ is increasing for $t > 0$ if $\tau(t) \leq e/t, t > 0$.

(ii) The condition $\tau(t) \leq e/t, t > 0$, is not necessary for the past entropy to be increasing. Indeed, if for instance $F(t) = t^c, 0 \leq t < 1, c > 0$, we have $\tau(t) = c/t, t > 0$, so that the reversed failure rate is decreasing for all $t > 0$, and thus Proposition (3.2.2) implies that $\bar{H}(t)$ is increasing in $t > 0$ even when $\tau(t) > e/t, i.e.$ when $c > e$.

(iii) For all positive t such that $\tau(t) \geq e/t$,the bound (3.2.7) is better than that given in (3.2.4).

The effect of increasing transformations on the monotonicity of residual entropy of random lifetimes has been considered in Ebrahimi and Kirmani [1996], where it is shown that, for all $t > 0$,

$$H_Y(t) = H_X\left(\phi^{-1}(t)\right) + E[log\phi'(X) \mid X > \phi^{-1}(t)],$$

where $Y = \phi(X)$, with ϕ strictly increasing, continuous, and differentiable.

Proposition 3.2.4 Let $Y = \phi(X)$, with ϕ strictly monotonic, continuous, and differentiable; then, for all $t > 0$,

$$\bar{H}_Y(t)$$
$$= \begin{cases} H_X\big(\phi^{-1}(t)\big) + E[log\{-\phi'(X)\}|X > \phi^{-1}(t)] & if\ \phi\ is\ strictly\ decreasing, \\ \bar{H}_X\big(\phi^{-1}(t)\big) + E[log\phi'(X)|X < \phi^{-1}(t)] & if\ \phi\ is\ strictly\ increasing. \end{cases}$$

If, in addition, ϕ is convex, then $\bar{H}_Y(t)$ is increasing in $t > 0$ if $H_X(t)$ is decreasing in $t > 0$ or if $\bar{H}_X(t)$ is increasing in $t > 0$.

Remark 3.2.3 When $Y = \phi(X)$ has the same distribution of X, Proposition (3.2.3) can be used to obtain various expressions involving H_X and $\bar{H}_X$.

(i) For instance, if X has a Pareto-type distribution, with $F(t) = (t/(1+t))\mathbf{1}_{(0,+\infty)}(t)$, then $Y = 1/X$ has the same distribution of X, so that

$$\bar{H}_X(t) = H_X\left(\frac{1}{t}\right) + E\left[log\{X^{-2}\} \mid X > \frac{1}{t}\right]$$
$$= log\frac{t}{t+1} + 2\left[1 - \frac{1}{t}log(1+t)\right].$$

(ii) If X is a random lifetime such that $Y = \beta - X$ has the same distribution of X then Proposition (3.2.3) yields $\bar{H}_X(t) = H_X(\beta - t), 0 < t < \beta$. For instance, such a relation holds if X has the following beta-type PDF:

$$f(x) = \frac{[x(\beta - x)]^{a-1}}{\beta^{2a-1}B(a,a)}, 0 < x < \beta,$$

with $a > 0, \beta > 0$, and $B(a,a) = \int_0^1 [x(1-x)]^{a-1}dx$ for $a > 0$.

Remark 3.2.4 Let $\phi_1(x) = \bar{F}_X(x)$ and $\phi_2(x) = F_X(x)$, with ϕ_1 and ϕ_2 satisfying the assumptions of Proposition (3.2.4). As $Y_1 = \phi_1(X)$ and $Y_2 = \phi_2(X)$ are uniformly distributed over $(0,1)$, recalling $(3.2.5)$ we have that, for all $t > 0$,

$$H_{U(0,1)}(t) = \begin{cases} H_X\left({\bar{F}_X}^{-1}(t)\right) + E\left[log\{-f_X(X)\}|X > {\bar{F}_X}^{-1}(t)\right], \\ \bar{H}_X\left({F_X}^{-1}(t)\right) + E\left[logf_X(X)|X < {F_X}^{-1}(t)\right]. \end{cases}$$

Remark 3.2.5 From Proposition (3.2.3) we have $\overline{H}_{aX}(t) = \overline{H}_X(t/a) + loga$ for all $t > 0$ and $a > 0$

3.3 Generalization of Past Entropy

Di Crescenzo and Longobardi [2002] introduced past entropy over $(0, t)$, since it is reasonable to presume that in many realistic situations uncertainty is not necessarily related to the future but can also refer to the past. If X denotes the lifetime of an item or of a living organism, then past entropy (or uncertainty of lifetime distribution) of an item is defined as

$$\overline{H}(X;t) = -\int_0^t \frac{f(x)}{F(t)} log\left(\frac{f(x)}{F(t)}\right) dx \qquad (3.3.1)$$

Gupta and Nanda [2002] generalized uncertainty of lifetime distribution, $H_1^{\beta^*}(X;t)$ and $H_2^{\beta^*}(X;t)$ by truncating the distributions above some point t as

$$H_1^{\beta^*}(X;t) = \frac{1}{\beta - 1}\left[1 - \int_0^t \left(\frac{f(x)}{F(t)}\right)^\beta dx\right] \qquad (3.3.2)$$

and

$$H_2^{\beta^*}(X;t) = \frac{1}{1 - \beta} log \int_0^t \left(\frac{f(x)}{F(t)}\right)^\beta dx \qquad (3.3.3)$$

Note that, as $\beta \to 1$, (3.3.2) and (3.3.3) reduce to $H^*(X;t)$ Again, $\beta \to 1$ and $t \to \infty$, (3.3.2) and (3.3.3) reduce to $H(X)$. $H_1^{\beta^*}(X;t)$ and $H_2^{\beta^*}(X;t)$ is known as first kind past entropy of order β and second kind past entropy of order β, respectively. One can easily verify that as $t \to \infty$, (3.3.2) reduces to $H_1(f)$ and (3.3.3) reduces to $H_2(f)$, defined in Gupta and Nanda [2002].

In forensic science, a lifetime distribution truncated above t is of utmost importance. Looking into this aspect, in this chapter, we analyze how the generalized past entropies behave when the distribution is truncated above t.

Based on generalized past entropy one stochastic order is defined and the properties are studied. It is shown that the stochastic order defined here is closed under increasing linear transformation. A nonparametric class is defined based on the generalized past entropy. The properties of the defined class are studied and some characterization results based on generalized past entropies are present. It is shown that, under certain condition, the generalized past entropies uniquely determine the distribution function. Here the uniform distribution is characterized in terms of generalized past entropies. Some discrete distribution results are also presented. Discrete uniform distribution is characterized in terms of generalized past entropy. Throughout this chapter, the words increasing and decreasing are not used in strict sense.

3.4 Properties of a Lifetime Random Variable Based on Generalized Past Entropy

Nanda and Paul $[2005]$ have defined an order of life distributions based on the measure $\bar{H}(X;t)$ as follows.

Definition 3.4.1 Let X and Y be two random variables, with support (l_X, u_X) and (l_Y, u_Y) denoting the lifetime of two components with density functions f and g, respectively. Then X is said to be greater than Y in past entropy order (written as $X \underset{\text{PE}}{\geq} Y$) if

$$\bar{H}(X;t) \leq \bar{H}(Y;t)$$

for all $t \in (max(l_X, l_Y), \infty)$. Here u_X and u_Y may be ∞ and l_X, l_Y may be zero.

Definition 3.4.2 Let X (resp. Y) be a random variable with support (l_X, u_X) (resp. (l_Y, u_Y)). Then X is said to be greater than Y in generalized past entropy of order β (written as $X \underset{\text{GPE}}{\geq} Y$) if

$H_1^{\beta^*}(X;t) \leq H_1^{\beta^*}(Y;t)$ (or equivalently, $H_2^{\beta^*}(X;t) \leq H_2^{\beta^*}(Y;t)$), for all $t \, \epsilon \, (max(l_X, l_Y), \infty)$. Here u_X and u_Y may be ∞ and l_X, l_Y may be zero.

Remark 3.4.1 It is to be noted that as $\beta \to 1$, Definitions 3.4.2 and 3.4.1 become identical.

Let $\bar{S}$ be the set of all pairs of distributions which are GPE ordered, for all $\beta > 0$ and S be the set of that which are PE ordered. Then clearly, $\bar{S} \supseteq S$. The following counterexample shows that $\bar{S}$ is a strict superset of S.

Counter example 3.4.1 Let X be a nonnegative random variable with distribution function

$$F(x) = \begin{cases} exp\left(-\frac{1}{2} - \frac{1}{x}\right) & if \ 0 \leq x \leq 1, \\ exp\left(-2 + \frac{x^2}{2}\right) & if \ 1 \leq x \leq 2, \\ 1 & if \ x \geq 2, \end{cases}$$

and Y be another nonnegative random variable with distribution function

$$G(x) = \begin{cases} \frac{x^2}{4} & if \ \ 0 \leq x \leq 2, \\ 1 & if \ x \geq 2 \end{cases}$$

Case 1: When $0 < t < 1$,

$$\int_0^t \frac{f(x)}{F(t)} log\left(\frac{f(x)}{F(t)}\right) dx - \int_0^t \frac{g(x)}{G(t)} log\left(\frac{g(x)}{G(t)}\right) dx$$

$$= \frac{1}{t} - logt + e^{1/t} \int_0^t \left(\frac{2}{x} - \frac{1}{x^3}\right) e^{-1/x} dx - log2 + \tfrac{1}{2} = B(t) \ \ (say)$$

Note that $B(0.0246) = -0.591694$ and $B(0.0248) = 7.744852$. Therefore, it is clear that $B(t)$ crosses the horizontal axis at some $t \, \epsilon \, (0,1)$. Thus, X and Y are not PE ordered. Again, note that

$$\int_0^t \left(\frac{f(x)}{F(t)}\right)^\beta dx - \int_0^t \left(\frac{g(x)}{G(t)}\right)^\beta dx = e^{\beta/t}\int_0^t \frac{1}{x^{2\beta}} e^{-\beta/x} dx - \frac{2^\beta}{\beta+1} t^{1-\beta}.$$

Case 2: When $1 < t < 2$,

$$\int_0^t \left(\frac{f(x)}{F(t)}\right)^\beta dx - \int_0^t \left(\frac{g(x)}{G(t)}\right)^\beta dx$$

$$= e^{-\beta t^2/2}\left[e^{3\beta/2}\int_0^1 \frac{1}{x^{2\beta}} e^{-\beta/x} dx + \int_1^t x^\beta e^{\beta x^2/2}\, dx\right] - \frac{2^\beta t^{1-\beta}}{\beta+1}.$$

Let us write

$$B_1(t) = \int_0^t \left(\frac{f(x)}{F(t)}\right)^\beta dx - \int_0^t \left(\frac{g(x)}{G(t)}\right)^\beta dx.$$

Then, it can be checked that

$B_1(t) =$

$$\begin{cases} e^{\beta/t}\displaystyle\int_0^t \frac{1}{x^{2\beta}} e^{-\beta/x} dx - \frac{2^\beta}{\beta+1} t^{1-\beta} & if\ 0 < t \le 1, \\ e^{-\beta t^2/2}\left[e^{3\beta/2}\displaystyle\int_0^1 \frac{1}{x^{2\beta}} e^{-\beta/x} dx + \int_1^t x^\beta e^{\beta x^2/2}\, dx\right] - \frac{2^\beta t^{1-\beta}}{\beta+1}, & if\ 1 \le \mathrm{t} < 2, \end{cases}$$

is negative for all values of $t \epsilon (0,2)$, when $\beta = 0.5$.

Case 3: When $t > 2$,

$$\int_0^t \left(\frac{f(x)}{F(t)}\right)^\beta dx - \int_0^t \left(\frac{g(x)}{G(t)}\right)^\beta dx$$

$$= e^{-\beta/2}\int_0^1 \frac{1}{x^{2\beta}} e^{-\beta/x} dx + e^{-2\beta}\int_1^2 x^\beta e^{\beta x^2/2}\, dx - \frac{2}{\beta+1}$$

$$= -0.0570091, \text{ when } \beta = 0.5.$$

Thus, combining all the cases we get that, when $\beta = 0.5, X \geq Y$, but $X \overset{PE}{\ngeq} Y$.

Ebrahimi and Pellerey [1995] mentioned for residual entropy that no relations exist between the orders based on residual entropy and the classical stochastic orders. Here we justify a similar claim by giving two examples.

The following example shows that $X \overset{ST}{\geq} Y$, but $X \overset{GPE}{\ngeq} Y$. A nonnegative random variable X is said to be greater than another nonnegative random variable Y in stochastic order (written as $X \overset{ST}{\geq} Y$ if $F(x) \leq G(x)$, for all $x > 0$, where F and G are the distribution functions of X and Y, respectively.

Example 3.4.1 Let X and Y be two nonnegative random variables having distribution functions as defined in Counterexample 3.4.1 Then, by writing $B_2(t) = F(t) - G(t)$, we have

$$B_2(t) = \begin{cases} exp\left(-\frac{1}{2} - \frac{1}{t}\right) - \frac{t^2}{4} & if\ 0 < t \leq 1, \\ exp\left(-2 + \frac{t^2}{2}\right) - \frac{t^2}{4} & if\ 1 \leq t \leq 2, \end{cases}$$

$\leq 0.$

Thus $X \overset{ST}{\geq} Y$. Again, we have, for $\beta = 2, B_1(1.6) = -0.0191663$ and $B_1(1.7) = 0.0333548$, where $B_1(t)$ is as defined in Counterexample 3.4.1 Hence, $X \overset{GPE}{\ngeq} Y$, for $\beta = 2$.

Below is an example which shows that $X \overset{ST}{\nleq} Y$, but $X \overset{GPE}{\leq} Y$.

Example 3.4.2 Let X be a nonnegative random variable having distribution function

$$F(x) = \begin{cases} \frac{x^2 + x}{4} & if\ 0 \leq x \leq 1, \\ \frac{x}{2} & if\ 1 \leq x \leq 2, \\ 1 & if\ x \geq 2. \end{cases}$$

and Y be another nonnegative random variable with distribution function

$$G(x) = \begin{cases} exp\left(\frac{1}{2} - \frac{1}{x}\right) & if \;\; 0 \le x \le 2, \\ 1 & if \; x \ge 2. \end{cases}$$

Then, $B_3(t) = F(t) - G(t)$, is given by

$$B_3(t) = \begin{cases} \frac{t^2 + t}{4} - exp\left(\frac{1}{2} - \frac{1}{t}\right) & if \; 0 \le t \le 1, \\ \frac{t}{2} - exp\left(\frac{1}{2} - \frac{1}{t}\right) & if \; 1 \le t \le 2. \end{cases}$$

and $B_4(t) = \int_0^t (f(x)/F(t))^\beta dx - \int_0^t (g(x)/G(t))^\beta dx$, is given as

$$B_4(t) = \begin{cases} \frac{(2t+1)^{\beta+1} - 1}{2(\beta+1)(t^2+t)^\beta} - e^{\beta/t} \int\limits_0^t \frac{1}{x^{2\beta}} e^{-\beta/x} dx & if \; 0 \le t \le 1, \\ \frac{3^{\beta+1} - 1}{2^{\beta+1}(\beta+1)t^\beta} + \frac{t-1}{t^\beta} - e^{\beta/t} \int\limits_0^t \frac{1}{x^{2\beta}} e^{-\beta/x} dx & if \; 1 \le \mathrm{t} \le 2. \end{cases}$$

Note that $B_3(0.3) = 0.03868352$ and $B_3(0.5) = -0.03563016$. Thus, it is clear that $B_3(t)$ crosses the horizontal axis. Therefore, $X \not\le^{ST} Y$. It can also be checked that $B_4(t)$ is negative for all $t \,\epsilon\, (0,2)$ when $\beta = 2$. Hence, $X \not\le_{GPE} Y$, for $\beta = 2$.

Proposition 3.4.1 For any absolutely continuous random variable X, define $Z = aX + b$, where $a > 0$ and $b \ge 0$ are constants. Then, for $t > b$,

$$(i) \;\; H_1^{\beta^*}(Z;t) = \frac{1}{\beta - 1}\left[1 - \frac{1}{a^{\beta-1}}\left\{1 - (\beta - 1)\, H_1^{\beta^*}\left(X; \frac{t-b}{a}\right)\right\}\right];$$

$$(ii) \;\; H_2^{\beta^*}(Z;t) = Ioga + \; H_2^{\beta^*}\left(X; \tfrac{t-b}{a}\right).$$

The following theorem shows that GPE order defined earlier is closed under increasing linear transformation.

Theorem 3.4.1 For two absolutely continuous random variables X and Y, define $Z_1 = a_1 X + b_1$ and $Z_2 = a_2 Y + b_2, a_1, a_2 > 0$ and $b_1, b_2 \ge 0$. Let

GPE GPE

$(i)\ X \geq Y, (ii)\ a_1 \geq a_2, (iii)\ b_1 \geq b_2$. Then $Z_1 \geq Z_2$, if $H_1^{\beta^*}(X;t)$ or $H_1^{\beta^*}(Y;t)$ is increasing in $t > b_1$.

Proof Suppose $H_1^{\beta^*}(X;t)$ is increasing in t. Since

$$(t-b_1)/a_1 \leq (t-b_2)/a_2,$$

we have

$$H_1^{\beta^*}\left(X;\tfrac{t-b_1}{a_1}\right) \leq H_1^{\beta^*}\left(X;\tfrac{t-b_2}{a_2}\right) \tag{3.4.1}$$

Further, $X \overset{\text{GPE}}{\geq} Y$ implies

$$H_1^{\beta^*}\left(X;\frac{t-b_2}{a_2}\right) \leq H_1^{\beta^*}\left(Y;\frac{t-b_2}{a_2}\right) \tag{3.4.2}$$

Combining (3.4.1) and (3.4.2) using Proposition 3.4.1, we have, $Z_1 \geq Z_2$.

GPE

If $H_1^{\beta^*}(Y;t)$ is increasing in t, the proof is similar and hence omitted.

The following example shows that $H_1^{\beta^*}(X;t)$ and $H_2^{\beta^*}(X;t)$ are increasing in $t > 0$.

Example 3.4.3 Let X be a nonnegative random variable having distribution function $F(x)$ as defined in Example 3.4.2. Then, we have

$$H_1^{\beta^*}(X;t) = \begin{cases} \dfrac{1}{\beta-1}\left[1-\dfrac{(2t+1)^{\beta+1}-1}{2(\beta+1)(t^2+t)^{\beta}}\right] & if\ 0 < t \leq 1, \\ \dfrac{1}{\beta-1}\left[1-\dfrac{3^{\beta+1}-1}{2^{\beta+1}(\beta+1)t^{\beta}}-\dfrac{t-1}{t^{\beta}}\right] & if\ 1 \leq \mathrm{t} \leq 2, \\ \dfrac{1}{\beta-1}\left[1-\dfrac{3^{\beta+1}-1}{(\beta+1)2^{2\beta+1}}-\dfrac{1}{2^{\beta}}\right] & if\ t \geq 2. \end{cases}$$

and

$$H_2^{\beta^*}(X;t)=\begin{cases}\dfrac{1}{1-\beta}log\left[\dfrac{(2t+1)^{\beta+1}-1}{2(\beta+1)(t^2+t)^\beta}\right] & if\ 0<t\le 1,\\ \dfrac{1}{1-\beta}log\left[\dfrac{3^{\beta+1}-1}{2^{\beta+1}(\beta+1)t^\beta}+\dfrac{t-1}{t^\beta}\right] & if\ 1\le \mathrm{t}\le 2,\\ \dfrac{1}{1-\beta}log\left[\dfrac{3^{\beta+1}-1}{(\beta+1)2^{2\beta+1}}+\dfrac{1}{2^\beta}\right] & if\ t\ge 2.\end{cases}$$

It is not very hard to check that, for $\beta=2$, $H_1^{\beta^*}(X;t)$ and $H_2^{\beta^*}(X;t)$ are increasing in t.

From the above example, we see that the class of distributions for which $H_1^{\beta^*}(X;t)$ (or equivalently $H_2^{\beta^*}(X;t)$) is increasing is not void.

Corollary 3.4.1 Let $Z_1=a_1X+b_1$ and $Z_2=a_2X+b_2, a_1, a_2>0$ and $b_1, b_2\ge 0$. Further, let $a_1\ge a_2$ and $b_1\ge b_2$.Then $Z_1 \overset{\text{PE}}{\ge} Z_2$ if $X \overset{\text{PE}}{\ge} Y$ and either $\bar{H}(X;t)$ or $\bar{H}(Y;t)$ is increasing in $t>b$.

Corollary 3.4.2 Let X and Y be two absolutely continuous random variables such that $X \overset{\text{GPE}}{\ge} Y$. Define, $X_1=aX+b$ and $Y_1=aY+b$, where $a>0$ and $b\ge 0$ are constants. Then, $X_1 \overset{\text{GPE}}{\ge} Y_1$ if either $H_1^{\beta^*}(X;t)$ or $H_1^{\beta^*}(Y;t)$ is increasing in $t\ge b$.

3.5 Nonparametric Class Based on Generalized Past Entropy

Nanda and Paul [2005] have defined one nonparametric class of life distributions based on the measure $\bar{H}(X;t)$ as follows.

Definition 3.5.1 A random variable X is said to have increasing uncertainty of life (IUL) if $\bar{H}(X;t)$ is increasing in $t\ge 0$.

Definition 3.5.2 A nonnegative random variable X is said to have IUL of order $\beta\big(IUL(\beta)\big)$ if $H_1^{\beta^*}(X;t)$ (or equivalently $H_2^{\beta^*}(X;t)$) is increasing in $t\ge 0$.

Remark 3.5.1 It can be noted that as $\beta\to 1$, Definitions 3.5.2 and 3.5.1 become identical.

The following counterexample shows that $IUL(\beta)$ class does not coincide, in general, with IUL class.

Counterexample 3.5.1 Let X be a nonnegative random variable as defined in Counterexample 3.4.1 When $0 < t < 1$,

$$L(t) \stackrel{\text{def}}{=} \int_0^t \frac{f(x)}{F(t)} log\left(\frac{f(x)}{F(t)}\right) dx = \frac{1}{t} - 2logt + e^{1/t} \int_0^t \left(\frac{2}{x} - \frac{1}{x^3}\right) e^{-1/x} dx$$

is not monotone, since $L(0.02) = 57.824$, $L(0.0243) = -34.1333$ and $L(0.03) = 6.09096$. On the other hand, by writing $L_1(t) = \int_0^t (f(x)/F(t))^\beta dx$ we have

$$L_1(t) = \begin{cases} e^{\beta/t} \int_0^t \frac{1}{x^{2\beta}} e^{-\beta/x} dx & if\ 0 < t \leq 1, \\ e^{-\beta t^2/2} \left[e^{3\beta/2} \int_0^1 \frac{1}{x^{2\beta}} e^{-\beta/x} dx + \int_1^t x^\beta e^{\beta x^2/2}\, dx \right] & if\ 1 \leq \mathrm{t} \leq 2, \\ 1.27632 & if\ t \geq 2. \end{cases}$$

which is increasing in t, for $\beta = 0.5$ Thus, we get that X is not IUL, but it is $IUL(0.5)$.

The following example prove that not all distributions are monotone in terms of $H_1^{\beta^*}(X; t)$.

Example 3.5.1 Let X be a nonnegative random variable with distribution function as defined in Counterexample 3.4.1 Then, for $1 \leq t \leq 2$,

$$\int_0^t \left(\frac{f(x)}{F(t)}\right)^\beta dx = e^{-\beta t^2/2} \left[e^{3\beta/2} \int_0^1 \frac{1}{x^{2\beta}} e^{-\beta/x} dx + \int_1^t x^\beta e^{\beta x^2/2}\, dx \right].$$

Hence, one can see that, for $\beta = 2$ and $1 \leq t \leq 2$,

$$H_1^{\beta^*}(X;t) = 1 - e^{-t^2}\left[e^3\int_0^1 \frac{1}{x^4}e^{-2/x}dx + \int_1^t x^2 e^{x^2}dx\right],$$

which is not monotone.

The following theorem shows that the nonparametric class given in Definition 3.5.2 is closed under linear transformation.

Theorem 3.5.1 Let $X \epsilon\, IUL(\beta)$. Define $Z = aX + b, a > 0$ and $b \geq 0$. Then $Z \epsilon\, IUL(\beta)$.

Proof The proof follows from the definition along with Proposition 3.4.1.

An application of the above theorem is given below.

Example 3.5.2 Let X be an exponentially distributed random variable having distribution function

$$F(x) = 1 - e^{-\lambda x}, x \geq 0, \lambda > 0,$$

and $Z = aX + b, a > 0$ and $b \geq 0$. Let us take $\beta = 2$. Then one can show that

$$H_1^{\beta^*}(X;t) = 1 - \int_0^t \left(\frac{f(x)}{F(t)}\right)^2 dx$$

$$= 1 - \frac{\lambda}{2}\left(\frac{1 + e^{-\lambda t}}{1 - e^{-\lambda t}}\right)$$

is increasing in $t > 0$. This implies that $X \epsilon\, IUL(2)$. Therefore, by using Theorem 3.5.1, we have $Z \epsilon\, IUL(2)$.

It is noted in Nanda and Paul [2005] that $DRHR \subset IUL$. Further, $IUL(\beta)$ reduces to IUL as $\beta \to 1$. Thus, $DRHR \subset IUL \subset IUL(\beta)$. A class of distributions is said to be $DRHR$ (decreasing in reversed hazard rate) if $\mu(x) = f(x)/F(x)$ is decreasing in x.

3.6 Characterization Results

In this section, we give a few characterizations of distributions in terms of generalized uncertainty.

Differentiating both sides of $(3.3.2)$ with respect to t, we get

$$\mu^{\beta}(t) - \beta\mu(t)\left[1 - (\beta - 1)\, H_1^{\beta^*}(X;t)\right] = -(\beta - 1)\frac{d}{dt}\, H_1^{\beta^*}(X;t),$$

where $\mu(t) = f(t)/F(t)$ is the reversed hazard rate function of the random variable X. Hence, for a fixed $t > 0, \mu(t)$ is a solution of $h(x) = 0$, where

$$\begin{aligned} h(x) = x^{\beta} - \beta x\left[1 - (\beta - 1)\, H_1^{\beta^*}(X;t)\right] \\ +(\beta - 1)\frac{d}{dt}\, H_1^{\beta^*}(X;t) \end{aligned} \tag{3.6.1}$$

Differentiating both sides of $(3.6.1)$ with respect to x, we gets

$$h'(x) = \beta x^{\beta-1} - \beta\left[1 - (\beta - 1)\, H_1^{\beta^*}(X;t)\right].$$

Note that, $h'(x) = 0$ gives $x = \left[1 - (\beta - 1)\, H_1^{\beta^*}(X;t)\right]^{1/(\beta-1)} = x_0$ (say).

Proposition 3.6.1 If $H_1^{\beta^*}(X;t)$ is increasing in $t > 0$, then

(i) $h(x) = 0$ has a unique solution if $h(x_0) = 0$. That unique solution is the reversed hazard rate.

(ii) $h(x) = 0$ has two solutions if $h(x_0) \neq 0$. Of these two solutions, at least one should be reversed hazard rate.

Proof We prove the theorem in two different cases.

Case 1: Let $\beta > 1$. Then $h(0) > 0$ since $H_1^{\beta^*}(X;t)$ is increasing in $t > 0$. Further, one can show that $h(x)$ is a convex function with minimum occurring at $x = x_0$. So, $h(x) = 0$ has unique solution when $h(x_0) = 0$.

Case 2: Let $\beta < 1$. Then $h(0) < 0$ and $h(x)$ is a concave function with maximum at $x = x_0$. So $h(x) = 0$ has unique solution when $h(x_0) = 0$.

Therefore, combining both the cases, we get that if $H_1^{\beta^*}(X;t)$ is increasing in $t > 0$, and $h(x_0) = 0$, then $h(x) = 0$ has the unique solution. Since, $\mu(t)$ is a solution to $h(x) = 0$, hence the unique solution is the reversed hazard rate.

The proof of (ii) can easily be obtained from the above two cases. If $h(x_0) \neq 0$, then from the above cases one can see that $h(x) = 0$ has two solutions and out of these two, one should be reversed hazard rate.

Below is one counterexample where both the solutions of $h(x) = 0$ are reversed hazard rates.

Counterexample 3.6.1 Let X be a nonnegative random variable having beta distribution with density function $f(t) = ct^{c-1}, 0 < t < 1$ and $= 0$, otherwise, where $\frac{1}{2} < c < 1$. Then it can be verified that, for $\beta = 2$, $H_1^{\beta^*}(X;t) = 1 - c^2/\big((2c-1)t\big)$, which is increasing in t, and $h(x_0) \neq 0$. Further, by writing $\mu(t)$ as the reversed hazard rate function of X, we have

$$\frac{x_0}{\mu(t)} = \frac{c}{2c-1} > 1$$

for $t \,\epsilon\, (0,1)$. Thus, for every $t > 0$, $h(x) = 0$ has two solutions $\mu(t)$ and $\mu^*(t)$ such that $\mu(t) < t_0 < \mu^*(t)$. Hence $\mu^*(t)$ must be a reversed hazard rate function.

Below we characterize the uniform distribution by the generalized uncertainty.

Theorem 3.6.1 The uniform distribution over $(a,b), a < b$ can be characterized by first kind uncertainty $H_1^{\beta^*}(X;t) = (\beta-1)^{-1}\left[1-(t-a)^{1-\beta}\right], a < t < b, i.e.$ a random variable X over $(a,b), a < b$ has uniform distribution if and only if $H_1^{\beta^*}(X;t) = (\beta-1)^{-1}\left[1-(t-a)^{1-\beta}\right], a < t < b.$

Proof The only if part is straight forward. To prove if part note that $H_1^{\beta^*}(X;t) = (\beta-1)^{-1}\left[1-(t-a)^{1-\beta}\right]$ gives

$$\int_0^t \left(\frac{f(x)}{F(t)}\right)^{\beta} dx = (t-a)^{1-\beta}, \quad a < t < b.$$

Differentiating both sides of the above expression with respect to t, we get

$$-\beta\mu(t)\int_0^t \left(\frac{f(x)}{F(t)}\right)^{\beta} dx + \mu^{\beta}(t) = (1-\beta)(t-a)^{-\beta},$$

or equivalently,

$\mu^{\beta}(t) - \beta\mu(t)(t-a)^{1-\beta} - (1-\beta)(t-a)^{-\beta} = 0.$

Thus, for a fixed $t > a, \mu(t)$ is a solution of $k(x) = 0$, where

$$k(x) = x^{\beta} - \beta x(t-a)^{1-\beta} - (1-\beta)(t-a)^{-\beta}. \qquad (3.6.2)$$

Differentiating both sides of $(3.6.2)$ with respect to x, we get

$$k'(x) = \beta x^{\beta-1} - \beta(t-a)^{1-\beta}.$$

Note that, $k'(x) = 0$ gives $x = (t-a)^{-1} = \bar{x}_0$, (say). Also, observe that $k(\bar{x}_0) = 0$.

Case 1: Let $\beta > 1$. Then, from $(3.6.2)$, we get $k(0) > 0$ and $k(x)$ is a convex function with minimum occurring at $x = \bar{x}_0$. So, $k(x) = 0$ has unique solution $x = \bar{x}_0$, since $k(\bar{x}_0) = 0$.

Case 2: Let $0 < \beta < 1$. Then $k(0) < 0$ and $k(x)$ is a concave function with maximum occurring at $x = \bar{x}_0$. Therefore, $k(x) = 0$ has unique solution $x = \bar{x}_0$.

Combining both the cases, we get that $k(x) = 0$ has the unique solution $x = \bar{x}_0$. Since, $\mu(t)$ is a solution to $k(x) = 0$, hence $\mu(t) = \bar{x}_0 = (t - a)^{-1}$. This is the reversed hazard rate function of the uniform distribution over $(a, b), a < b$. Hence the theorem is established.

Chapter - IV

Some New Results on Cumulative Residual Entropy

4.1 Introduction

Shannon [1948] proposed a measure of uncertainty in a discrete distribution based on the Boltzmam entropy of classical statistical mechanics. He called it the entropy. The Shannon entropy of a discrete random variable X is defined by

$$H(X) = - \sum_i p_i log p_i \qquad (4.1.1)$$

where $p_i^{'}$s are the probabilities of the random variable X.

With this he opened up a new branch of mathematics with far reaching applications in many areas. To name a few: Information Theory (Cover and Thomas [1991]) Statistics (Kullback [1959]) Financial Analysis (Sharpe [1985]) and Data Compression (Salomon [1998]) etc.

This measure of uncertainty has many important properties which agree with our intuitive notion of randomness. We mention three: (1) It is always positive. (2) It vanishes if and only if it is a certain event. (3) Entropy is increased by the addition of an independent component, and decreased by conditioning.

However, extension of this notion to continuous distribution poses some challenges. A straight forward extension of the discrete case to continuous case X with density f called differential entropy reads as:

$$H(X) = -\int f(x) log f(x) dx \tag{4.1.2}$$

This definition raises the following concerns:

1) It is only defined for distributions with densities. For e.g., there is no definition of entropy for a mixture density comprising of a combination of Gaussians and delta functions.
2) The entropy of a discrete distribution is always positive, while the differential entropy of a continuous variable may take any value on the extended real line.
3) It is "inconsistent" in the sense that the differential entropy of a uniform distribution in an interval of length a is $log\ a$, which is zero if $a = 1$, negative if $a < 1$, and positive if $a > 1$.
4) The entropy of a discrete distribution and the differential entropy of a continuous variable are decreased by conditioning. Moreover, if X and Y are discrete (continuous) random variables, and the conditional entropy (differntial entropy) of X given Y equals the entropy (differntial entropy) of X, then X and Y are independent. Also, the conditional entropy of the discrete variable X given Y is zero, if and only if X is a function of Y, but the vanishing of the conditional differential entropy of X given Y does not imply that X is a function of Y.
5) Use of empirical distributions in approximations is of great value in practical applications. However. It is impossible, in general, to approximate the differential entropy of a continuous variable using the entropy of empirical distributions.
6) Consider the following situation: Suppose X and Y are two discrete random variables, with X taking on values $\{1,2,3,4,5,6\}$, each with a probability $1/6$ and Y taking on values $\{1,2,3,4,5,6\}$ again each with probability $1/6$. The information content measured in these two

random variables using Shannon entropy is the same $i.e.$, Shannon entropy does not bring out any differences between these two cases. However, if the two random variables represented distinct pay off schemes in a game of chance, the information content in the two random variables would be considered as being dramatically different. Nevertheless Shannon entropy fails to make any distinction whatsoever between them Jumarie [1990].

In this chapter we present an alternative measure of uncertainty in a random variable X and call it the Cumulative Residual Entropy (CRE) of X. The main objective of our study is to extend Shannon entropy to random variables with continuous distributions. The concept presented in this chapter overcomes the problems mentioned above, while retaining many of the important properties of Shannon entropy. For instance, both are decreased by conditioning, while increased by independent addition. They both obey the data processing inequality etc. However, the differential entropy doesn't have the following important properties of CRE.

1) CRE has consistent definitions in both the continuous and discrete domains;
2) CRE is always non-negative;
3) CRE can be easily computed from sample data and these computations asymptotically converge to the true values.
4) The conditional CRE (defined below in section 4.3) of X given Y is zero, if and only if X is a function of Y.

The basic idea is to replace the density function with the cumulative distribution in Shannon's definition (4.1.2). The distribution function is more regular than the density function, because the density is computed as the derivative of the distribution. Moreover, in practice what is of interest and/or measurable is the distribution function. For example, if the random variable is the life span of a machine, then the event of interest is not whether the life span equals t, but rather whether the life span exceeds t. Our definition also preserves the well established principle that the logarithm of the probability of an event should represent the information content in the event.

In section 4.2, we present the definition of cumulative residual entropy (CRE) and its properties. In section 4.3, we present the definition of the conditional CRE and its properties.

4.2. Cumulative Residual Entropy: A New Measure of Information

In this section, we define an alternate measure of uncertainty, $i.e.$ cumulative residual entropy (CRE) in a random variable and then derive some properties about this new measurement.

Rao et al. $[2004]$ introduced an alternative measure of uncertainly called the cumulative residual entropy of a non- negative random vector X defined as

$$\mathcal{E}(X) = -\int_0^{\infty} \bar{F}(x) log \bar{F}(x) dx \tag{4.2.1}$$

where $\bar{F}_X(t) = P(X > t)$ is the survival (reliability) function.

The measure $(4.2.1)$ is suitable to describe the information in problems related to the ageing properties in reliability theory $(\text{Rao } [2005] \text{ and Drissi et al.} [2008])$. The multivariate analog of $(4.2.1)$ is given by

$$\mathcal{E}(X) = -\int_{\mathcal{R}_+} P(|X| > \lambda) log P(|X| > \lambda) d\lambda \tag{4.2.2}$$

where X is a random vector $\mathcal{R}_+$. The various generalization of CRE corresponding to the generalization of Shannons entropy are presented here. Asadi and Zohravand $[2007]$ generalized the Rao et al. $[2004]$ cumulative residual entropy. It is defined as the cumulative residual entropy of the random variable $X_t = |X - t|\, X > t$ and is given by

$$\mathcal{E}(X,t) = -\int_t^{\infty} \frac{\bar{F}(x)}{\bar{F}(t)} log \frac{\bar{F}(x)}{\bar{F}(t)} dx \tag{4.2.3}$$

Kumar, V and Taneja, H.C. [2010] have proposed the cumulative entropy of order α, based on Rènyi's entropy defined as

$$\mathcal{E}_{\alpha}(X;t) = \frac{1}{1-\alpha} log \left[\frac{\int_t^{\infty} \bar{F}^{\alpha}(x)dx}{\bar{F}^{\alpha}(t)}\right] \tag{4.2.4}$$

and have characterized the residual lifetime distributions. The measure (4.2.4) has also been studied by Sunoj and Linu [2010].

Kumar, V and Taneja, H.C. [2010] further generalizes the Eq. (4.2.1) and called it cumulative residual entropy of order α and type β based on Verma's entropy of the random variable X as

$$\mathcal{E}_{\alpha}^{\beta}(X) = \frac{1}{\beta-\alpha} log \left[\int_0^{\infty} \bar{F}^{\alpha+\beta-1}(x)dx\,,\right], \qquad \beta-1<\alpha<\beta, \beta \geq 1 \tag{4.2.5}$$

when $\beta = 1$, $\alpha \to 1$ (4.2.5) reduces to

$$\lim_{\beta=1,\alpha\to 1} \mathcal{E}_{\alpha}^{\beta}(X) = -\int_0^{\infty} \bar{F}(x) log \bar{F}(x)dx,$$

which is the cumulative residual entropy (4.2.1) as suggested by Rao et al. [2004].

The cumulative residual entropy of order α and type β of the random variable X with survival function $\bar{F}_t(x)$ is proposed as

$$\mathcal{E}_{\alpha}^{\beta}(X) = \frac{1}{\beta-\alpha} log \left[\int_t^{\infty} \bar{F}_t^{\;\alpha+\beta-1}(x)dx\right] \tag{4.2.6}$$

This can be rewritten as

$$\mathcal{E}_{\alpha}^{\beta}(X;t) = \frac{1}{\beta-\alpha} log\left[\frac{\int_t^{\infty}\bar{F}^{\alpha+\beta-1}(x)dx}{\bar{F}^{\alpha+\beta-1}(t)}\right] \qquad (4.2.7)$$

When $\beta = 1$, then $(4.2.7)$ reduces to $(4.2.4)$, a cumulative entropy of order α, given by Kumar et al. [2010].

Example 4.2.1 (CRE of the uniform distribution)

Consider a general uniform distribution with the density function:

$$p(x) = \begin{cases} \frac{1}{a} & 0 \le x \le a \\ 0 & o.w \end{cases} \qquad (4.2.8)$$

Then, its CRE is computed as follows

$$\mathcal{E}(X) = -\int_0^a P(X > x)\, logP(X > x)dx$$

$$= -\int_0^a \left(1 - \frac{x}{a}\right) log\left(1 - \frac{x}{a}\right) dx$$

$$= \frac{1}{4}a \qquad (4.2.9)$$

Example 4.2.2 (CRE of the exponential distribution)

The exponential distribution with mean $1/\lambda$ has the density function:

$$p(x) = \lambda e^{-\lambda x} \qquad (4.2.10)$$

Correspondingly, the CRE of the exponential distribution is:

$$\mathcal{E}(X) = -\int_0^{\infty} e^{-\lambda x}\, log\, e^{-\lambda x} dt$$

$$= \int_0^{\infty} \lambda t\, e^{-\lambda x} dt$$

$$= \frac{1}{\lambda} \qquad (4.2.11)$$

Example 4.2.3 (CRE of the gaussian distribution)

The Gaussian probability density function is:

$$p(x) \; = \; \frac{1}{\sqrt{2\pi}\sigma} exp\left[-\frac{(x-m)^2}{2\sigma^2}\right], \qquad (4.2.12)$$

where m is the mean and σ^2 is the variance.

The cumulative distribution function is:

$$F(x) = 1 - erfc\left(\frac{x-m}{\sigma}\right), \qquad (4.2.13)$$

where $erfc$ is the error function:

$$erfc(x) \; = \; \frac{1}{\sqrt{2\pi}} \int_x^{\infty} exp(-t^2/2)dt.$$

Then the CRE of the Gaussian distribution is:

$$\mathcal{E}(X) \; = \; -\int_0^{\infty} erfc\left(\frac{x-m}{\sigma}\right) log\left[erfc\left(\frac{x-m}{\sigma}\right)\right] dx \qquad (4.2.14)$$

Properties of CRE

Using the convexity of $xlogx$, it is easy to see CRE is a concave function of distribution.

Theorem 4.2.1 $\mathcal{E}(X) < \infty$ if for all i and some $p > N, E[\,X_i\,|\,^p] < \infty$.

Proof For simplicity we write the proof in steps

Step 1: Using H$\ddot{o}$lders inequality

$$E\left[\prod_{i=1}^{N} f_i\right] \le \prod_{i=1}^{N} E[f_i^N]^{1/N}$$

we see that for sets $A_1, \ldots, A_n$

$$P[A_1 \cap A_2 \cap \ldots \cap A_N] \le E\left[\prod_{i=1}^{N} 1_{A_i}\right] \le \prod_{i=1}^{N} P(A_i)^{1/N} \qquad (4.2.15)$$

Step 2: It is not difficult to see that for each $0 \le \alpha \le 1$

$$|xlogx| \le e^{-1}\frac{x^{\alpha}}{1-\alpha} \quad 0 \le x \le 1 \tag{4.2.16}$$

Step 3: From step 2 and step 1, for any $0 < \alpha < 1$ and any i,

$$P\big[|X_i| > \ x_{i,} \ 1 \le i \le N\big]\big|\, logP\big[|X_i| > x_{i,} 1 \le i \le N\big]\big|$$

$$\le \frac{e^{-1}}{1-\alpha} P\big[|X_i| > x_{i,} 1 \le i \le N\big]^{\alpha}$$

$$\le \frac{e^{-1}}{1-\alpha} \prod_{i=1}^{N} P[|X_i| > x_i]^{\alpha/N} \tag{4.2.17}$$

Integrating both sides of (4.2.17), on $\mathcal{R}_{+}^{N}$, we get

$$\mathcal{E}(X) \ \le \ \frac{e^{-1}}{1-\alpha}\int_{\mathcal{R}_{+}^{N}} \prod_{i=1}^{N} P[|X_i| > x_i]^{\alpha/N}\, dx$$

$$= \frac{e^{-1}}{1-\alpha} \prod_{i=1}^{N} \Big\{\int_0^{\infty} P[|X_i| > x_i]^{\alpha/N}\, dx_i\Big\} \tag{4.2.18}$$

Step 4: We have for any positive random variable Y

$$\int_0^{\infty} P[\,Y > y\,]^{\alpha/N} dy = \int_0^{1} P[\,Y > y\,]^{\alpha/N} dy + \int_1^{\infty} P[\,Y > y\,]^{\alpha/N} dy$$

$$\le 1 + \int_1^{\infty} \left\{\frac{1}{y^p} E[\,Y^p]\right\}^{\alpha/N} dy \tag{4.2.19}$$

The second inequality on the right hand side above follows from the Markov inequality. The last integral is finite if $\frac{p\alpha}{N} > 1$, i.e. if $p < \frac{N}{\alpha}$. For $p > N$, we can choose $\alpha < 1$ to satisfy $p > \frac{N}{\alpha}$. Then the conclusion of the theorem follows from (4.2.18) and (4.2.19).

The traditional Shannon entropy of a sum of independent variables is larger than that of either, as shown by the following theorem:

Theorem 4.2.2 For any non-negative and independent variables X and Y,

$$max\big(\mathcal{E}(X), \mathcal{E}(Y)\big) \ \le \ \mathcal{E}(X+Y)$$

Proof Since X and Y are independent

$$P[\, X+Y>t\,] \;=\int dF_Y\,(a) P[\, X>t-a\,],$$

where F_Y is the cumulative distribution function of Y. Using Jensen's inequality,

$$\begin{aligned} &P[X+Y>t] log\, P[X+Y>t] \\ &\qquad\qquad \leq \int dF_Y\,(a) P[\, X>t-a\,] log P[\, X>t-a\,] \end{aligned}$$

Integrating both sides with respect to t from 0 to ∞,

$$\begin{aligned} -\mathcal{E}(X+Y) \;\leq\; &\int dF_Y\,(a) \int_0^\infty P[X>t-a] log P[X>t-a] dt \\ = &\int dF_Y\,(a) \int_0^\infty P[X>t-a] log P[X>t-a] dt \\ = -&\int dF_Y\,(a) \mathcal{E}(X) = \; -\mathcal{E}(X) \end{aligned}$$

where in the first equality we used that for $t\; \leq a\; P[X>t-a]=1$, and in the second one we change variables in the inner integral.

Next theorem shows one of the salient features of CRE, in the discrete case, Shannon entropy is always non-negative, and equals zero if and only if the random variable is a certain event. However, this is not valid for the Shannon entropy in the continuous case as defined in $(4.1.1)$. In contrast, in this regard CRE does not differentiate between discrete and continuous cases, as shown by the following theorem:

Theorem 4.2.3 $\mathcal{E}(X)\;\geq 0$ and equality holds if and only if

$P[|X|=\;\lambda]=1$ for some vector λ, i.e. $|X_i|=\;\lambda_i$ with probability 1.

Proof Now $xlogx=0$ if and only if $x=0$ or 1. Thus $\mathcal{E}(X)=0$ implies $P[|X|>\lambda]=0$ or 1 for almost all λ. If for all λ, $P[|X|>\lambda]=0$, then $P[|X|=0]=1$. Now we consider the case that for some λ, $P[|X|>\lambda]=1$.

Note that if λ and μ satisfy $P[|X| > \lambda] = P[|X| > \mu] = 1$, then also $P[|X| > \lambda \vee \mu] = 1$, where $\lambda \vee \mu$ denotes the vector whose coordinates are maxima of coordinates of λ and μ. Then,

$$\lambda_c = max_{\lambda\,\epsilon}\ \lambda,$$

satisfies $P[|X| = \lambda_c] = 1$, where $\wedge = \{\lambda | P[|X| > \lambda] = 1\}$, and the maximum of $\lambda\,\epsilon\wedge$ is a vector whose coordinates are the maxima of the coordinates of all $\lambda\,\epsilon\wedge$.

Proposition 4.2.1 If X and Y are two non-negative random variables with finite means $E(X)$ and $E(Y)$, respectively, and such that $X \leq_{st} Y$, then

$$\mathcal{E}(X) \leq \mathcal{E}(Y) - E(X)log\frac{E(X)}{E(Y)}. \qquad (4.2.20)$$

Proof Using the log-sum inequality we have

$$\int_0^\infty \bar{F}(x)log\frac{\bar{F}(x)}{\bar{G}(x)}dx \geq \int_0^\infty \bar{F}(x)dx\ log\frac{\int_0^\infty \bar{F}(x)dx}{\int_0^\infty \bar{G}(x)dx}.$$

$$= E(X)log\frac{E(X)}{E(Y)}$$

Hence we obtain

$$\mathcal{E}(X) = \int_0^\infty \bar{F}(x)log\bar{F}(x)dx$$

$$\leq -\int_0^\infty \bar{F}(x)log\bar{G}(x)dx - E(X)log\frac{E(X)}{E(Y)} \qquad (4.2.21)$$

Finally, using that $\bar{F} \leq \bar{G}$, we obtain

$$\mathcal{E}(X) \leq -\int_0^\infty \bar{F}(x)log\bar{G}(x)dx - E(X)log\frac{E(X)}{E(Y)}$$

$$\leq -\int_0^{\infty} \bar{G}(x) log \bar{G}(x) dx - E(X) log \frac{E(X)}{E(Y)}$$

$$= \mathcal{E}(Y) - E(X) log \frac{E(X)}{E(Y)}$$

Now using the Weibull distribution we obtain an upper bound for the CRE similar to that obtained by Rao et al. [2004].

Proposition 4.2.2 If X is a non-negative random variable, then

$$\mathcal{E}(X) \leq \frac{E\left(X^{\beta+1}\right) \Gamma^{\beta} \left(1 + \frac{1}{\beta}\right)}{(\beta + 1) E^{\beta}(X)} \tag{4.2.22}$$

for all $\beta > 0$.

Proof If Y is a random variable with a Weibull distribution and reliability function $\bar{G}(t) = e^{-(\lambda t)^{\beta}}$, from (4.2.21), we obtain

$$-\mathcal{E}(X) \geq \int_0^{\infty} \bar{F}(x) log \bar{G}(x) dx + E(X) log \frac{E(X)}{E(Y)}$$

$$= E(X) log \frac{E(X)}{E(Y)} - \int_0^{\infty} (\lambda x)^{\beta} \bar{F}(x) dx$$

$$= E(X) log \frac{E(X)}{E(Y)} - \lambda^{\beta} \frac{E\left(X^{\beta+1}\right)}{\beta + 1},$$

where

$$\mu = E(Y) = \int_0^{\infty} e^{-(\lambda x)^{\beta}} dx = \frac{\Gamma\left(1 + \frac{1}{\beta}\right)}{\lambda}.$$

Hence

$$-\mathcal{E}(X) \geq E(X) log \frac{E(X)}{\mu} - \frac{E\left(X^{\beta+1}\right) \Gamma^{\beta} \left(1 + \frac{1}{\beta}\right)}{(\beta + 1)} \mu^{-\beta},$$

which is maximized for a fixed β at

$$\mu_\beta = \left(\frac{\beta E\left(X^{\beta+1}\right)\Gamma^\beta\left(1+\frac{1}{\beta}\right)}{(\beta+1)E(X)}\right)^{1/\beta}.$$

Substituting this value we obtain

$$-\mathcal{E}(X) \geq E(X) log\frac{E(X)}{\mu_\beta} - \frac{E\left(X^{\beta+1}\right)\Gamma^\beta\left(1+\frac{1}{\beta}\right)}{(\beta+1)}\mu_\beta^{-\beta}$$

$$= -\frac{E(X)}{\beta} log\left(\frac{\beta E\left(X^{\beta+1}\right)\Gamma^\beta\left(1+\frac{1}{\beta}\right)}{(\beta+1)E^{\beta+1}(X)}\right) - \frac{E(X)}{\beta},$$

and using that $-logx \geq 1-x$, we have

$$-\mathcal{E}(X) \geq \frac{E(X)}{\beta}\left(1 - \frac{\beta E\left(X^{\beta+1}\right)\Gamma^\beta\left(1+\frac{1}{\beta}\right)}{(\beta+1)E^{\beta+1}(X)}\right) - \frac{E(X)}{\beta}$$

$$= -\frac{E\left(X^{\beta+1}\right)\Gamma^\beta\left(1+\frac{1}{\beta}\right)}{(\beta+1)E^{\beta}(X)}.$$

In particular, if we take $\beta = 1$, then we obtain expression similar to that obtained by Rao et al. [2004].

4.3 Conditional Cumulative Residual Entropy

Definition 4.3.1 Given a random $\mathcal{R}^N$ vector X and a $\sigma-$ field $\mathcal{F}$, we define the conditional CRE : $\mathcal{E}(X|\mathcal{F})$ by

$$\mathcal{E}(X|\mathcal{F}) = -\int_{\mathcal{R}_+} P(|X| > x|\mathcal{F}) logP(|X| > x|\mathcal{F})dx \qquad (4.3.1)$$

where, $P(|X| > x|\mathcal{F})$ denotes the conditional expectation of the indicator function namely, $E\left(\boldsymbol{I}_{(|X|>x)}/\mathcal{F}\right)$. Note that $\mathcal{E}(X|\mathcal{F})$ is a random variable measurable with respect to $\mathcal{F}$. For example, if $\mathcal{F}$ is the $\sigma-$field generated by a random variable Y then, $\mathcal{E}(X/\mathcal{F}) = g(Y)$ where, $g(Y) = -\int P\left(|X| > \frac{x}{Y} = y\right) log\,(P(|X| > x/Y = y))dx$. When $\mathcal{F}$ is the trivial field

$$\mathcal{E}(X|\mathcal{F}) = \mathcal{E}(X)$$

The following proposition says that the conditional CRE has the "super-martingale property".

Proposition 4.3.1 Let $X \in L^p$ for some $p > N,$ then for $\sigma -$ field $\mathcal{G} \subset \mathcal{F}$

$$E[\mathcal{E}(X|\mathcal{F})|\mathcal{G}] \leq \mathcal{E}(X|\mathcal{G}) \qquad (4.3.2)$$

Proof The proof follows by applying Jensen's inequality $(1.7.1)$ for the convex function $xlogx$.

In words, Proposition 4.3.1 states that conditioning decreases CRE. The same result holds for the Shannon Entropy. In particular

$$H(Z|X,Y) \leq H(Z|X),$$

for any random variables X, Y and Z. A simple consequence is the Data Processing Inequality.

If $X \to Y \to Z$ is a Markov chain, $i.e.$, if the conditional distribution of Z given (X, Y) equals that given Y, then $H(Z|Y) \leq H(Z|X)$. This is so because $H(Z|X,Y) = H(Z|Y)$, from Markov property.

For the same reason we have the data processing inequality for CRE:

If $X \to Y \to Z$ is Markov,

$$E[\mathcal{E}(Z|Y)] \leq E[\mathcal{E}(Z|X)].$$

Indeed, $\mathcal{E}(Z|Y) = \mathcal{E}(Z|X,Y)$, and from (4.3.2)

$$E[\mathcal{E}(Z|X,Y)|X] \leq \mathcal{E}(Z|X)$$

Theorem 4.3.1 For any X and $\sigma -$field $\mathcal{F}$

$$E[\mathcal{E}(X|\mathcal{F})] \leq \mathcal{E}(X). \tag{4.3.3}$$

Equality holds iff X is independent of $\mathcal{F}$. Where, X is said to be independent of the σ $-$field $\mathcal{F}$ means X is independent of every random variable measurable with respect to $\mathcal{F}$.

Proof The inequality $(4.3.3)$ follows from $(4.3.2)$ by taking $\mathcal{G}$ to be the trivial field. Now we prove the necessary and sufficient condition for the equality. First it is clear from the definition that if $|X|$ is independent of $\mathcal{F}$ equality holds in $(4.3.2)$. Conversely, suppose that there is equality in $(4.3.3)$. By Jensen's inequality for conditional expectations,

$$E\big[P[|X| > x|\mathcal{F}]logP[|X|> x|\mathcal{F}]\big]$$

$$\geq P[|X| > x]logP[|X| > x] \tag{4.3.4}$$

for all x. Integrating both sides of $(4.3.4)$ with respect to x, and using $(4.3.3)$ we see that equality holds in $(4.3.4)$ for almost all x. Note the following fact, which can be proved using Taylor with remainder:

Fact: *For any random variable X and strictly convex* $\varphi(i.e.\varphi > 0)$

$$E[\varphi(X)] = \varphi(E[X])$$

implies $X = E(X)$ almost surely.

Using this fact and strict convexity of $xlogx$, we get from $(4.3.4)$

$$P[|X| > x|\mathcal{F}] = P[|X| > x]$$

for $a.e.\ x.\ i.e.|X|$ is independent of $\mathcal{F}$.

4.4 Characterization of Cumulative Residual Entropy

Definition 4.4.1 A random variable X is said to be increasing (decreasing) $DCRE$, denoted by $IDCRE$ $(DDCRE)$, if $\mathcal{E}(X;t)$ is an increasing (decreasing) function of t.

Definition 4.4.2 A random variable X is said to be increasing (decreasing) failure rate, denoted by IFR (DFR), if $r_X(t)$ is increasing (decreasing) in t.

Definition 4.4.3 A random variable X is said to be increasing (decreasing) mean residual life, denoted by $IMRL\,(DMRL)$, if $e_X(t)$ is increasing (decreasing) in t.

Asadi and Zohrevand $[2007]$ obtained characterizations for the exponential, power and Pareto distributions from the following relationship between the $DCRE$ and the MRL

$\mathcal{E}(X;t) = ce_X(t),$

where c is a non-negative real constant. In the following theorem, we extend this result to the more general case where c is a function of t.

Theorem 4.4.1 Let X be a non-negative absolutely continuous random variable such that $\mathcal{E}(X;t) = c(t)e_X(t)$ for $t \geq 0$, then

$$e_X(t) = \left(K - \int_0^t \big(1 - c(x)\big)e^{c(x)}dx\right)e^{-c(t)}, \tag{4.4.1}$$

with $K = \mu e^{c(0)}$ and $\mu = E(X)$.

Proof If X has survival function $\bar{F}_X(t)$, from (4.2.3) we have

$$\mathcal{E}(X;t)\bar{F}_X(t) = log\bar{F}_X(t)\int_t^\infty \bar{F}_X(x)dx - \int_t^\infty \bar{F}_X(x)log\bar{F}_X(x)dx.$$

Differentiating with respect to t, we obtain

$$\mathcal{E}'(X;t) = r_X(t)\big(\mathcal{E}(X;t) - e_X(t)\big), \tag{4.4.2}$$

We know that the relationship between $r_X(t)$ and $e_X(t)$ is

$$r_X(t) = \frac{e_X'(t) + 1}{e_X(t)} \tag{4.4.3}$$

which jointly with $\mathcal{E}(X;t) = c(t)e_X(t)$ and (4.4.3) give

$$e'_X(t) + c'(t)e_X(t) = c(t) - 1$$

Solving this linear differential equation we obtain (4.4.1).

In the next example we show how to use this general result to obtain new characterization results.

Example 4.4.1 If $c(t) = at + b$ for $t > 0$ and $a > 0$ from Theorem 4.4.1, we obtain the general model with mean residual life function

$$e_X(t) = \frac{b - 2 + at}{a} - \frac{(b - 2)e^{-at}}{a} + Ke^{-at-b}.$$

If $a = 0$, from Theorem 4.4.1, we obtain the characterization results given by Asadi and Zohrevand [2007] for the exponential, power and Pareto distributions. Other characterization results can be obtained from our general result by solving the corresponding differential equation.

The following examples show that the *only if* part of Asadi and Zohrevand [2007] is not necessarily true, that is, $IDCRE$ is not equivalent to $IMRE$ and $DDCRE$ is not equivalent to $DMRL$. Note that this fact make the new classes $IDCRE$ and $DDCRE$ more interesting since now they are not equal to the known classes $IMRE$ and $DMRL$.

Example 4.4.2 If X has a Burr type XII distribution with parameters $c = 1.5$ and $k = 2$, the survival function of X is given by

$$\bar{F}_X(t) = \left(1 + t^{3/2}\right)^{-2} \text{ for } t > 0.$$

The corresponding mean residual life function of X, $e_X(t)$, has a minimum at $t = 0.269462$ (see Fig. 1). Hence, the function $e_X(t)$ is decreasing to the minimum and it is increasing later. Therefore, X is not $IMRE$. Moreover $e'_X(0) = -1$. However, it can be checked that the $DCRE$ at time t of X is an increasing function in t, that is, X is $IDCRE$ (see Fig. 1).

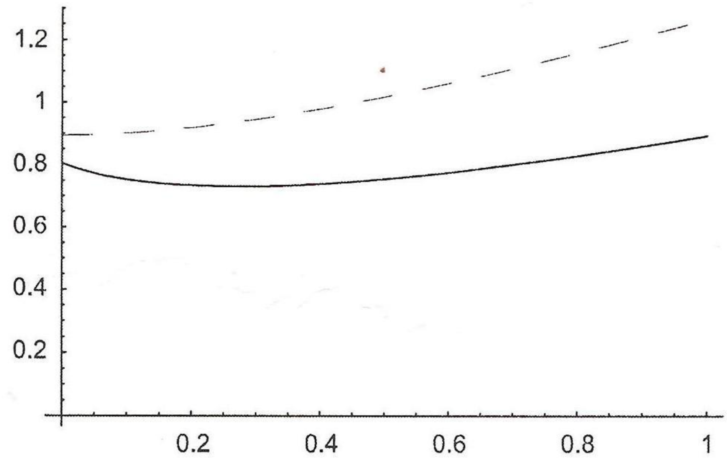

Fig. 1: MRL (Continuous line) and DCRE (dashed line) functions of the Burr type XII distribution given in Example 4.4.2

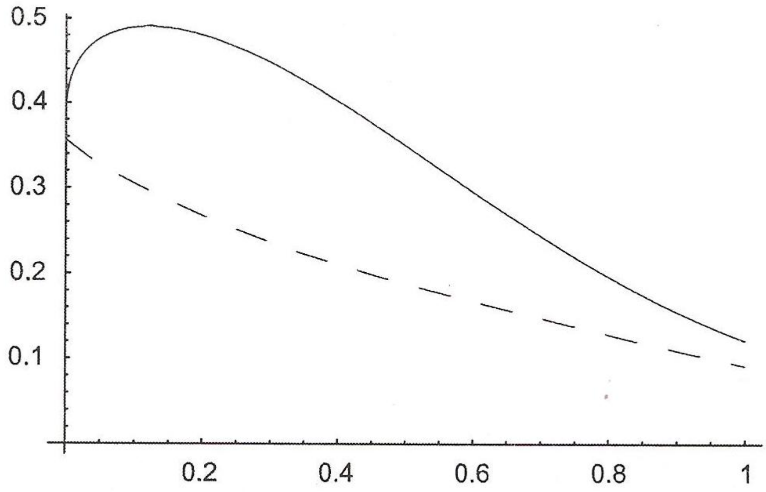

Fig. 2: MRL (continuous line) and DCRE (dashed line) functions of the survival function given in Example 4.4.3

The following example shows that $DDCRE$ does not imply $DMRL$.

Example 4.4.3 Let X be a continuous random variable with survival function

$$\bar{F}_X(t) = e^{-t^5-(2t)^{1/2}} \quad \text{for } t \geq 0$$

The mean residual life function $e_X(t)$ of X has a maximum at $t = 0.120786$ (see Fig. 2). Hence, the function $e_X(t)$ is increasing to the maximum and it is decreasing later. Therefore, X is not $DMRL$. However, it can be checked that the $DCRE$ at time t of X is a decreasing function in t, that is, X is $DDCRE$ (Fig. 2).

Remark 4.4.1 Note that by the well known implications between IFR (DFR) and $DMRL$ $(IMRL)$ classes, Asadi and Zohrevand $[2007]$, the following implications hold:

$$IFR\ (DFR)\ \Rightarrow DMRL\ (IMRL)\ \Rightarrow DDCRE\ (IDCRE).$$

Now, we also know that the reverse implications are not necessarily true. Thus, the $DDCRE$ property can be seen as a necessary property for the properties IFR and $DMRL$.

Example 4.4.4 Let X and Y be two random variables with survival functions

$$\bar{F}_X(t) = \begin{cases} 1 - 0.5t & for\ 0 \leq t \leq 1, \\ 0.5e^{-(t-1)} & for\ t > 1, \end{cases}$$

and

$$\bar{F}_Y(t) = \begin{cases} 1 - 0.5t^2 & for\ 0 \leq t \leq 1, \\ 0.5e^{-(t-1)} & for\ t > 1. \end{cases}$$

The failure rate functions of X and Y are given by

$$r_X(t) = \begin{cases} \dfrac{0.5}{1 - 0.5t} & for\ 0 \leq t \leq 1, \\ 1 & for\ t > 1, \end{cases}$$

and

$$r_Y(t) = \begin{cases} \dfrac{t}{1 - 0.5t^2} & for\ 0 \leq t \leq 1, \\ 1 & for\ t > 1, \end{cases}$$

Note that $r_X(t) > r_Y(t)$ in $t\ \epsilon\ [0, 0.585786)$ and $r_X(t) < r_Y(t)$ in $t\ \epsilon\ (0.585786, 1]$.

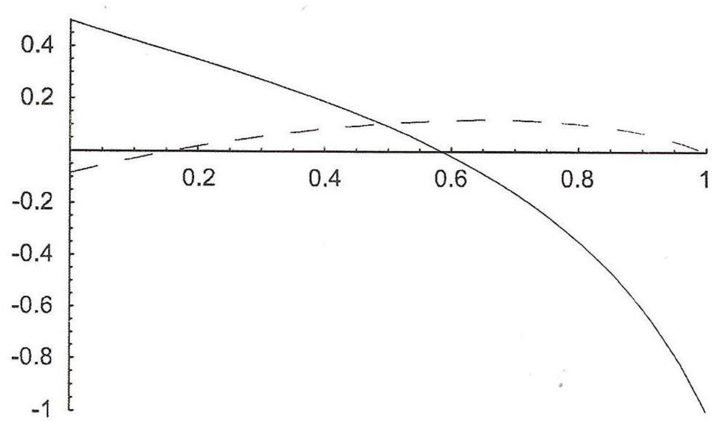

Fig. 3: Difference of failure rate functions (continuous line) and mean residual life functions (dashed line) of the survival functions given in Example 4.4.4

The mean residual life functions of X and Y are given by

$$e_X(t) = \begin{cases} \dfrac{1.25 - t + 0.25t^2}{1 - 0.5t} & for\ 0 \le t \le 1, \\ 1 & for\ t > 1\,, \end{cases}$$

and

$$e_Y(t) = \begin{cases} \dfrac{\frac{4}{3} - t + \frac{1}{6}t^3}{1 - 0.5t^2} & for\ 0 \le t \le 1, \\ 1 & for\ t > 1\,. \end{cases}$$

Note that $e_X(t) < e_Y(t)$ in $t \epsilon\, [0, 0.154534)$ and $e_X(t) > e_Y(t)$ in $t \epsilon\, (0.154534, 1]$. Fig. 3 shows the difference of failure rate functions $\left(r_X(t) - r_Y(t)\right)$ and the difference of mean residual life functions $\left(e_X(t) - e_Y(t)\right)$. Thus we have that, for all $t \epsilon\, (0.154534, 0.585786)$, $r_X(t) > r_Y(t)$ and $e_X(t) > e_Y(t)$.

Proposition 4.4.1 Let X be a non-negative continuous random variable and let Y be the equilibrium random variable associated to X. Then,

$$H(Y; t) = \log e_X(t) + \frac{\mathcal{E}(X; t)}{e_X(t)}.$$

The proof is easy. Using this relationship we obtain the following result. First we need a lemma taken from Asadi and Zohrevand $[2007]$. The proof is immediate from (4.4.1).

Lemma 4.4.1 *X is $DDCRE$ $(IDDCRE)$ if and only if,*

$$\mathcal{E}(X;t) \leq e_X(t)\ (\geq)\ \text{for all } t.$$

Proposition 4.4.2 *Let X be* a non-negative continuous random variable and let Y be the equilibrium random variable associated to X. Then

1. $X\ DDCRE \Leftrightarrow Y\ DURL$

2. $X\ IDCRE \Leftrightarrow Y\ IURL$

Proof From Proposition $(4.4.1)$, we obtain

$$H'(Y;t) = \frac{e'_X(t)}{e_X(t)} + \frac{\mathcal{E}'(X;t)e_X(t) - \mathcal{E}(X;t)e'_X(t)}{e_X^2(t)}.$$

Using now $(4.4.2)$ we get

$$H'(Y;t) = \frac{e'_X(t)}{e_X(t)} + \frac{r_X(t)e_X(t)\big(\mathcal{E}(X;t) - e_X(t)\big) - \mathcal{E}(X;t)e'_X(t)}{e_X^2(t)}.$$

Then, a straightforward calculation gives

$$H'(Y;t) = \frac{r_X(t)e_X(t) - e'_X(t)}{e_X^2(t)}\big(\mathcal{E}(X;t) - e_X(t)\big).$$

Hence using Lemma 4.5.1 the proof is complete.

Proposition 4.4.3 Let X be a non-negative continuous random variable and let Y be the equilibrium random variable associated to X. Then, if X is $IMRL$, then $\mathcal{E}(X;t) \leq \mathcal{E}(Y;t)$ for all t. In particular, $\mathcal{E}(X) \leq \mathcal{E}(Y)$.

Proof It is well known that if Y is the equilibrium random variable associated to X then, $X \leq_{fr} Y$ if, and only if, X is $IMRL$. Moreover, X is $IMRL$, if, and only if, Y is DFR. Hence, Y is $IMRL$. The proof is completed Asadi and Zohrevand $[2007]$.

The proportional odds family (POF) also known as tilt parameter family is a semi-parametric family useful in the study of survival and reliability data (e.g., Marshall and Olkin, $[2007]$). If F is a distribution function and $\bar{F} = 1 - F$, the POF associated to F is defined by the reliability function

$$\bar{F}(t|p) = \frac{p\bar{F}(t)}{1-(1-p)\bar{F}(t)}, \qquad (4.4.4)$$

for $p > 0$. For this family we obtain the following result.

Proposition 4.4.4 Let X be a non-negative continuous random variable and let X_p be a random variable with reliability defined by $(4.4.4)$. If $0 < p < 1$ $(p > 1)$ and X (resp $.X_p$) is $IMRL$, then $\mathcal{E}(X;t) \geq \mathcal{E}(Y;t)$ $(\leq)$ for all t.

Proof If r is the failure rate of X, then the failure rate of X_p is given by

$$r(t|p) = \frac{1}{1-(1-p)\bar{F}(t)} r(t).$$

Then $X \geq_{fr} X_p$ $\left(\leq_{fr}\right)$ if, and only if, $0 < p < 1$ $(p > 1)$. Hence the proof is completed Asadi and Zohrevand $[2007]$.

Chapter - V

Generalized Residual Entropy of Order Statistics and its Applications

5.1 Introduction

The aim of the present chapter is to study properties of Rényi's Residual Entropy (RRE) of order statistics and record values. Let $X_1, X_2, \dots, X_n$ denote a random sample of size n from F. Order statistics refer to the arrangement of $X_1, X_2, \dots, X_n$ from the smallest to the largest, denoted as $X_{1:n} \le X_{2:n} \le \cdots \le X_{n:n}$. Order statistics are used in many branches of probability and statistics including characterization of probability distributions, goodness-of-fit tests, quality control, reliability analysis and many other problems. Order statistics are of particular interest in reliability theory in the study of the lifetime properties of the coherent systems and in life testing, when data are collected based on different censoring mechanisms, Daivid and Nagaraja [2003].

A record value of a sequence of independent identically distributed $(\mathrm{i.i.d})$ random variables $\{X_i; i \ge 1\}$ is an observation X_j whose value exceeds or is less than the values of all previous observations. For example, X_j is an upper record if $X_i < X_j$ for every $i < j$. Record values arise naturally in problems such as industrial stress testing, meteorological analysis, hydrology, sporting and athletic events, and economics. In reliability, records model are used to study, for example, technical systems which are subject to shocks, e.g. peaks of voltages. Successive large shocks may be viewed as realizations of records from a sequence of identically independent voltages, Arnold et al. [1998] and Nevzorov [2001].

Several authors have studied the information properties of ordered data. Wong and Chen $[1990]$ showed that the difference between the average entropy of order statistics and the entropy of the parent distribution is a constant. They also showed that when the distribution of the data is symmetric, the entropy of order statistics is symmetric about the median. Park $[1995]$ obtained some recurrence relations for the entropy of order statistics. Ebrahimi et al. $[2004]$ explored some properties of the Shannon entropy of the order statistics and showed that the Kullback–Leibler information functions involving order statistics are distribution free. Baratpour et al. $[2007, 2008]$ obtained some results for the Shannon entropy and Rényi entropy of the order statistics and record values.

In Section 5.2, we present the RRE of order statistics $X_{k:n}$ of a sample from any continuous distribution function F in terms of RRE of order statistics $X_{k:n}$ of a sample from uniform distribution. Since for many statistical models the functional form of the RRE of order statistics cannot be obtained in a closed form, we obtain upper and lower bounds for RRE of order statistics. Several illustrative examples are given. We also show that, under some mild conditions, the $RRE's$ of the minimum and maximum of a random sample are monotone functions of the number of observations of sample. We give a counter example to show that the RRE of other order statistics $X_{k:n}$ is not necessary monotone function of n. We also study the monotone behavior of RRE of order statistics $X_{k:n}$ in terms of k. It is shown that the RRE of $X_{k:n}$ is not a monotone function of k on the entire support of F. In Section 5.3, we investigate properties of RRE of record values. We give bounds for RRE of record values. We also show that, under some mild conditions, the RRE of record values is monotone function of number of records in the sequence.

Before proceeding to give the main results of the chapter, we overview some preliminary concepts on partial orderings between random variables

Shaked and Shanthikumar [2007]. Let X and Y be two random variables with survival functions $\bar{F}$ and $\bar{G}$ and density functions f and g respectively.

Definition 5.1.1

(a) The random variable Y is said to be smaller than X in the usual stochastic order (denoted by $Y \leq_{\text{st}} X$) if $\bar{G}(x) \leq \bar{F}(x)$ for all x.

(b) The random variable Y is said to be smaller than X in likelihood ratio order (denoted by $Y \leq_{lr} X$) if $\frac{f(x)}{g(x)}$ is an increasing function of x.

It can be shown that if $Y \leq_{lr} X$, then $Y \leq_{\text{st}} X$ Shaked and Shanthikumar [2007].

Throughout the chapter increasing (decreasing) means non-decreasing (non-increasing).

5.2 The Residual Rényi Entropy of Order Statistics

In this section, we focus on the RRE of order statistics. First note that the density function and survival function of $X_{k:n}$ denoted by $f_{k:n}(x)$ and $\bar{F}_{k:n}(x)$, $k = 1, \dots, n$, respectively, are

$$f_{k:n}(x) = \frac{1}{B(k, n-k+1)} [F(x)]^{k-1} [1 - F(x)]^{n-k} f(x) \qquad (5.2.1)$$

and

$$\bar{F}_{k:n}(x) = \sum_{i=0}^{k-1} \binom{n}{i} F^{i}(x) \bar{F}^{n-i}(x), \qquad (5.2.2)$$

where

$$B(a, b) = \int_0^1 x^{a-1} (1-x)^{b-1} dx, \qquad a > 0, b > 0.$$

The survival function $\bar{F}_{k:n}(x)$ can also be represented as

$$\bar{F}_{k:n}(x) = \frac{\bar{B}_{F(x)}(k, n-k+1)}{B(k, n-k+1)}, \tag{5.2.3}$$

where

$$\bar{B}_x(a,b) = \int_x^1 u^{a-1}(1-u)^{b-1}\, du, \qquad 0 < x < 1.$$

$B(a,b)$ and $\bar{B}_x(a,b)$ are known as the beta and the incomplete beta functions, respectively David and Nagaraja [2003].

Notation: Throughout this section we use the notation $Y \sim \bar{B}_t(a,b)$ to show that Y has a truncated beta distribution with density function

$$f_Y(y) = \frac{1}{\bar{B}_t(a,b)} y^{a-1}(1-y)^{b-1}, \qquad t \le y \le 1. \tag{5.2.4}$$

Remark 5.2.1 In reliability engineering $(n-k+1)$-out-of-n systems are very important kind of structures. A $(n-k+1)$-out-of-n system functions if and only if at least $(n-k+1)$ components out of n components function. If $X_1, X_2, \dots, X_n$ denote the independent lifetimes of the components of such system, then the lifetime of the system is equal to the order statistics $X_{k:n}$. The special cases of $k=1$ and $k=n$ are corresponding to series and parallel systems, respectively. Assuming that a $(n-k+1)$-out-of-n system is working at time t, then the RRE of $X_{k:n}$ measures the entropy of the residual lifetime of the system. Hence the RRE, as a dynamic measure of entropy, can be important for system designers to get information about the entropy of used $(n-k+1)$-out-of-n systems at any time t.

The following lemma shows that the RRE of order statistics of uniform distribution can be written in terms of incomplete beta function which is important in computational point of view.

Lemma 5.2.1 Let $U_{k:n}$ be kth order statistic based on a random sample of size n from uniform distribution on $(0,1)$. Then

$$H_\alpha(U_{k:n};t) = \frac{1}{1-\alpha} log\bar{B}_t(\alpha(k-1)+1, \alpha(n-k)+1)$$

$$-\frac{\alpha}{1-\alpha} log\bar{B}_t(k, n-k+1).$$

If F is a continuous distribution function then it is well known, from the probability integral transformation, that

$$U_{k:n} \stackrel{d}{=} F(X_{k:n})\ k = 1, \dots, n,$$

where d stands for equality in distribution and $X_{k:n}$ is the kth order statistic based on a random sample of size n from F David and Nagaraja $[2003]$. Using this, in the following theorem, we will show that the RRE of order statistics $X_{k:n}$ can be represented in terms of RRE of order statistics of uniform distribution.

Theorem 5.2.1 Let F be an absolutely continuous distribution function with density f. Then the RRE of the kth order statistic can be represented in terms of the RRE of kth order statistic from uniform distribution, over the unit interval, as follows:

$$H_\alpha(X_{k:n};t) = H_\alpha\big(U_{k:n}; F(t)\big) + \frac{1}{1-\alpha} logE\left[f^{\alpha-1}\big(F^{-1}(Y_k)\big)\right] \qquad (5.2.5)$$

where $Y_k \sim \bar{B}_{F(t)}(\alpha(k-1)+1, \alpha(n-k)+1)$.

It can be seen, after some calculations, that when in $(5.2.5)$ $\alpha \to 1$, the Shannon entropy of kth order statistic from a sample of F can be written as follows:

$$H(X_{k:n};t) = H\big(U_{k:n}; F(t)\big) - E\left[logf\big(F^{-1}(Y_k)\big)\right] \qquad (5.2.6)$$

where $Y_k \sim \bar{B}_{F(t)}(k, n-k+1)$. The specialized version of this result for $t = 0$, was already obtained by Ebrahimi et al. $[2004]$.

Remark 5.2.2 The quantity $f\left(F^{-1}(x)\right)$ is known, in the literature, as the density-quantile function and is used to approximate the moment of order statistics David and Nagaraja $[2003]$.

In the following we give some examples.

Example 5.2.1 Suppose that F is exponential with mean $\frac{1}{\theta}$. Then $f\left(F^{-1}(y)\right) = \theta(1-y)$ and we have

$$E\left[f^{\alpha-1}\left(F^{-1}(Y_1)\right)\right] = \frac{\theta^{\alpha-1}e^{\theta t(1-\alpha)}}{n\alpha[\alpha(n-1)+1]}.$$

For $k = 1$, Theorem 5.2.1 gives

$$H_\alpha(X_{1:n};t) = \frac{log\alpha}{\alpha-1} - log(n\theta).$$

On the other hand, we have

$$H_\alpha(X;t) = \frac{log\alpha}{\alpha-1} - log\theta.$$

This gives

$$H_\alpha(X_{1:n};t) - H_\alpha(X;t) = -logn.$$

That is, in the exponential case the difference between of RRE of the lifetime of a series system and RRE of the lifetime of each components is free of both of the time and α depends only on the number of components of the system.

Example 5.2.2 Let X have Pareto distribution with distribution function

$$F(x) = 1 - \left(\frac{\beta}{x}\right)^\theta, \qquad x \geq \beta > 0, \theta > 0$$

and density function

$$f(x) = \frac{\theta\beta^\theta}{x^{\theta+1}}, \qquad x \geq \beta > 0, \theta > 0.$$

Then

$$f\left(F^{-1}(y)\right) = \frac{\theta}{\beta}(1-y)^{1+\frac{1}{\theta}}.$$

Therefore, for the first order statistic of a random sample of size n we have

$$E\left[f^{\alpha-1}\left(F^{-1}(Y_1)\right)\right] = \left[\frac{\alpha(n-1)+1}{\alpha(n\theta+1)-1}\right]\theta^{\alpha}\beta^{1-\alpha}\left(\frac{\beta}{t}\right)^{(\alpha-1)(1+\theta)}.$$

Hence Theorem 5.2.1 gives

$$H_\alpha(X_{1:n};t) = logt + \frac{\alpha}{1-\alpha}logn\theta - \frac{1}{1-\alpha}log[\alpha(n\theta+1)-1],$$

$$\alpha > \frac{1}{n\theta+1}$$

The difference between RRE of $X_{1:\,n}$ and RRE of X is

$$H_\alpha(X_{1:n};t) - H_\alpha(X;t) = \frac{\alpha}{1-\alpha}log\theta - \frac{1}{1-\alpha}log\left[\frac{\alpha(n\theta+1)-1}{\alpha(\theta+1)-1}\right],$$

$$\alpha > \frac{1}{n\theta+1}$$

which is free of t and depends only on α and n.

We obtained the closed form of RRE of the first order statistics in exponential and Pareto distributions. However, we do not have a closed form for RRE of other order statistics for these distributions. This is true, in general for other distributions, that there is no closed form for the RRE of order statistics. This gives a motivation for obtaining some bounds for RRE of order statistics which is present in the following theorem.

Theorem 5.2.2 Let X be a non-negative continuous random variable with density function f and distribution function F. Let also $H_\alpha(X;t)$ and $H_\alpha(X_{k:n};t)$ denote the $RRE's$ of X and $X_{k:n}$, respectively.

(a) Let $H_\alpha(X;t)$ be finite. If $m_k = max\left\{F(t), \frac{k-1}{n-1}\right\}$, then for

$\alpha > 1\ (0 < \alpha < 1)$

$$H_\alpha(X_{k:n};t) \geq (\leq) b_k(t) + H_\alpha(X;t) + \frac{\alpha}{1-\alpha} log\bar{F}(t),$$

where

$$b_k(t) = \frac{\alpha}{1-\alpha}\big[(k-1)logm_k + (n-k)log(1-m_k)$$
$$- log\bar{B}_{F(t)}(k, n-k+1)\big].$$

Proof According to Theorem 5.2.1, it is enough to obtain a bound for $\frac{1}{1-\alpha}logE\left[f^{\alpha-1}\left(F^{-1}(Y_k)\right)\right]$. Note that $m_k = max\left\{F(t), \frac{k-1}{n-1}\right\}$ is the mode of the distribution of Y_k. Let $M_k = f_{Y_k}(m_k)$, then for $\alpha > 1$ $(0 < \alpha < 1)$ we have

$$\frac{1}{1-\alpha} logE\left[f^{\alpha-1}\left(F^{-1}(Y_k)\right)\right]$$

$$= \frac{1}{1-\alpha} log \int_{F(t)}^{1} \frac{y^{\alpha(k-1)}(1-y)^{\alpha(n-k)}}{\bar{B}_{F(t)}(\alpha(k-1)+1, \alpha(n-k)+1)} f^{\alpha-1}\left(F^{-1}(y)\right)dy$$

$$\geq (\leq) \frac{1}{1-\alpha} \log M_k + \frac{1}{1-\alpha} log \int_{F(t)}^{1} f^{\alpha-1}\left(F^{-1}(y)\right)dy$$

$$= \frac{1}{1-\alpha} logM_k + \frac{1}{1-\alpha} log \int_{t}^{\infty} f^{\alpha-1}(u)du$$

$$= \frac{1}{1-\alpha} logM_k + H_\alpha(X;t) + \frac{\alpha}{1-\alpha} log\bar{F}(t).$$

(b) Let $M = f(m) < \infty$, where $m = sup\{x: f(x) \leq M\}$ is mode of the distribution. Then for $\alpha > 0$

$$H_\alpha(X_{k:n};t) \geq H_\alpha\left(U_{k:n}; F(t)\right) - logM.$$

Proof We have for $\alpha > 1$ $(0 < \alpha < 1)$,

$$f^{\alpha-1}\left(F^{-1}(Y_k)\right) \leq (\geq) M^{\alpha-1}.$$

Thus for the RRE of kth order statistics, using Theorem 5.2.1, we can write

$$H_\alpha(X_{k:n};t) \geq H_\alpha\big(U_{k:n};F(t)\big) - logM \ .$$

For the case when $\alpha \to 1$, using relation $(5.2.6)$, we obtain the following result for residual Shannon entropy

$$H(X_{k:n};t) \geq H\big(U_{k:n};F(t)\big) - logM.$$

Part (a) of the theorem gives an lower bound, in the case of $\alpha > 1$ (upper bound, in the case of $0 < \alpha < 1$), for RRE of $X_{k:n}$ in terms of incomplete beta function and the RRE of the parent distribution. Part (b) of the theorem shows a lower bound for RRE of $X_{k:n}$ in terms of RRE of order statistics of uniform distribution and the mode of the underlying distribution. The bound in part (b) is particulary interesting since it shows that the difference between the RRE of $X_{k:n}$ at t and RRE of $U_{k:n}$ at $F(t)$ is atleast $-logM$, where $M = f(m)$ and m is the mode of X.

In Table1, we list the bounds of the RRE of the order statistics based on Theorem 5.2.2 for some well known distributions.

Table 1

Bounds for $H_\alpha(X_{k:n};t)$ based on Parts (a) and (b) of Theorem 5.2.2., respectively.

Density function	Bonds
Triangular distribution	
$f(x) = \begin{Bmatrix} \frac{2x}{\delta}, 0 \le x \le \delta & \\ & \partial > 0 \\ \frac{2(1-x)}{1-\delta}, \delta \le x \le 1, & \end{Bmatrix}$	$\ge (\le) \begin{Bmatrix} bk(t) + \frac{1}{1-\propto}\left[\propto log\frac{2}{\delta} - \log(1=\propto) + \log(\delta^x - t^{1+x}),\right] & 0 \le t \le \delta \\ bk(t) + \frac{1}{1-\propto}\left[\propto log2 + \log\frac{1-t}{1=\propto} + \propto \log\frac{1-t}{1-\propto}\right], & \delta \le t \le 1 \end{Bmatrix}$ $\le H \propto (U_{k;n}F(t)) - log2$
Standard half-Cauchy distribution	$\ge (\le) b_k(t) - log2 + \frac{1}{1-\propto}\left[-\propto log\pi + log\bar{B}_{\frac{t^2}{1+t}\ \ 2}\left(\frac{1}{2}, \propto - \frac{1}{2}\right)\right]$
$F(x) = \frac{2}{\pi(1+x^2)}, x \ge 0$	$\ge H_x\ (U_{k:n}F(t)) - log\frac{2}{\pi}$
Standard half-normal distribution	$\ge (\le) b_k(t) + \frac{1}{2}\log(2\pi\sigma^2) + \frac{1}{1-\propto}\left[log\overline{\varphi}\left(\frac{\sqrt{\propto}\ (t-\mu}{\propto}\right) + log\frac{2^x}{\sqrt{\propto}}\right]$
$f(x) = \frac{2}{\propto\sqrt{2\pi}}\exp(-\frac{(x-\mu)^2}{2\sigma^2}, x > \mu \ge 0$	$\ge H_\propto\ (U_{k:n}F(t)) + \frac{1}{2}\ log\frac{\pi\sigma^2}{2}$
Generalized exponential distribution	$\ge (\le) b_k(t) + log\theta + \frac{1}{1-\propto}\left[\propto log\beta + log\bar{B}_{1-exp}\left(\frac{\mu-t}{\theta}\right)(\propto(\beta-1) + 1, \propto\right]$
$f(x) = \frac{\beta}{\theta}\exp\left(\frac{\mu-x}{\theta}\right)(1-\exp\ (\frac{\mu-x}{\theta}))^\beta - 1,$ $x\mu > 0, \beta, \theta > 0.$	$\ge H_\propto\ (U_{k:n}F(t)) + log\theta + (1-\beta)\left(1-\frac{1}{\beta}\right), \beta > 1$
Generalized gamma distribution	$\ge (\le) b_k(t) - \frac{1}{c}\log(\propto\sigma) - logc + \frac{1}{1-\propto}\left[log\Gamma\left(\propto\frac{(ca-)+1}{c};\ \propto\sigma t^c\right)\propto\log(\propto^a\ \Gamma(a))\right]$
$F(x)\frac{c\sigma^a}{\Gamma(a)}X^{c1}e - \sigma x^{t,}\ x > 0, \alpha c, \sigma > 0$	$\ge H_\propto\left(U_{k:n}F(t)\right) + \log\left(\frac{\Gamma(\alpha)c^{\alpha-1}}{(\sigma c)^{\frac{1}{c}}}\right) - \left(\frac{c\alpha-1}{c}\right)\log(c\alpha-1) + \frac{(ca-1)}{c}, c\alpha > 1$

In the following we explore monotone behavior of RRE of order statistics. First we prove the following lemma which plays a crucial role in the subsequent results.

Lemma 5.2.2 Let $u(x)$ and $v_\lambda(x), \lambda > 0$, be non-negative functions where $u(x)$ is increasing. Assume that $0 \le t < c \le \infty$ and W_λ has a density function f_λ where

$$f_\lambda(w) = \frac{u^{m\lambda}(w)v_\lambda(w)}{\int_t^c u^{m\lambda}(x)v_\lambda(x)dx}, \qquad w \in (t,c) \qquad (5.2.7)$$

Let m be real valued and define function h_α as follows.

$$h_\alpha(m) = \frac{1}{1-\alpha} log \frac{\int_t^c u^{m\alpha}(x)v_\alpha(x)dx}{\left(\int_t^c u^m(x)v_1(x)dx\right)^\alpha}, \quad \alpha > 0, \alpha \neq 1 \qquad (5.2.8)$$

(i) If for $\alpha > 1 (0 < \alpha < 1), W_\alpha \le_{st} (\ge_{st}) W_1$ then $h_\alpha(m)$ is an increasing function of m.

(ii) If for $\alpha > 1 (0 < \alpha < 1), W_\alpha \ge_{st} (\le_{st}) W_1$ then $h_\alpha(m)$ is a decreasing function of m.

Proof We prove part (i). Part (ii) can be proved similarly. Under the assumption that $h_\alpha(m)$ is differentiable in terms of m, we obtain

$$\frac{\partial h_\alpha(m)}{\partial m} = \frac{1}{1-\alpha}\left(\frac{\frac{\partial g_\alpha(m)}{\partial m}}{g_\alpha(m)}\right)$$

where

$$g_\alpha(m) = \frac{\int_t^c u^{m\alpha}(x)v_\alpha(x)dx}{\left(\int_t^c u^m(x)v_1(x)dx\right)^\alpha}$$

and

$$\frac{\partial g_\alpha(m)}{\partial m}$$

$$= \frac{\alpha}{\left(\int_t^c u^m(x)v_1(x)dx\right)^{\alpha+1}}\left[\int_t^c logu(x)u^{m\alpha}(x)v_\alpha(x)dx\int_t^c u^m(x)v_1(x)dx\right.$$

$$\left.-\int_t^c logu(x)u^m(x)v_1(x)dx\int_t^c u^{m\alpha}(x)v_\alpha(x)dx\right]. \quad (5.2.9)$$

Now using the fact that $W_\alpha \leq_{st} (\geq_{st})W_1$ and that log is increasing function we get

$$E[logu(W_\alpha)] \leq (\geq)E[logu(W_1)]$$

Shaked and Shanthikumar [2007]. This shows that (5.2.9) is non positive (non- negative) and hence $h_\alpha(m)$ is an increasing function of m.

Remark 5.2.3 Under the assumptions of Lemma 5.2.2, it can also be proved that when $u(x)$is decreasing then

(a) For $\alpha > 1 (0 < \alpha < 1), W_\alpha \leq_{st} (\geq_{st})W_1$ implies that $h_\alpha(m)$ is a decreasing function of m.

(b) For $\alpha > 1 (0 < \alpha < 1), W_\alpha \geq_{st} (\leq_{st})W_1$ implies that $h_\alpha(m)$ is an increasing function of m.

Using Lemma 5.2.2, we can now prove the following corollary for $(n-k+1)$-out-of-n systems with components having uniform distributions.

Corollary 5.2.1 (a) Consider a parallel (series) system consists of n components where the components have uniform distribution over unit interval. Then the RRE of the system lifetime is a decreasing function of the number of components.

Proof We assume that the system is parallel. For series system similar arguments can be used to prove the result on using Remark 5.2.3. From Lemma

5.2.1 it is easily seen that $H_\alpha(U_{n:n};t)$ can be written as (5.2.8) with $u(x)=x$ and $v_\alpha(x)=x^{-\alpha}$ in which we assume, without loss of generality, $n\geq 1$ is a continuous variable. Since for $\alpha>1$ $(0<\alpha<1)$ the ratio

$$\frac{\int_t^1 x^{\alpha(n-1)}dx}{\int_t^1 x^{n-1}dx}$$

is increasing (decreasing) in t, for the chosen $u(x)$ and $v_\alpha(x)$, we have

$$W_\alpha \geq_{st} (\leq) W_1,$$

where density function of $W_\lambda, \lambda>0$, is given in (5.2.7). Hence from Lemma 5.2.2 we conclude that the RRE of the parallel system is a decreasing function of the number of components.

(b) Let $U_{k:n}$ denote the kth order statistic of uniform distribution over unit interval. If $k_1\leq k_2\leq n$ are integers then for $t\geq \frac{k_2-1}{n-1}$,

$$H_\alpha\left(U_{k_1:n};t\right)\leq H_\alpha\left(U_{k_2:n};t\right)$$

Proof The result can be proved using the same arguments as used to prove part (a) on taking $u(x)=\frac{x}{1-x}$ and $v_\alpha(x)=\frac{(1-x)^{n\alpha}}{x^\alpha}$. In this case it can be easily seen that for $t\geq\frac{k-1}{n-1}$ and $\alpha>1$ $(0<\alpha<1)$,

$$W_\alpha \leq_{st} (\geq) W_1,$$

Hence it can be concluded that for $k_1\leq k_2\leq n$,

$$H_\alpha\left(U_{k_1:n};t\right)\leq H_\alpha\left(U_{k_2:n};t\right),\quad t\geq\frac{k_2-1}{n-1} \tag{5.2.10}$$

The class of distribution functions with decreasing density functions is a wide class of distributions. Examples are exponential, Pareto, mixture of exponential and Pareto distributions, etc. There are also distribution functions with increasing density; for example the power distribution with density $f(x)=\beta x^{\beta-1}, 0<x<1, \beta>1$.

5.3 The Residual Rényi Entropy of Record Values

In this section we obtain some results on the RRE of record values. Let $U_1, U_2, \ldots$ be a sequence of upper record values based on a sequence of non-negative continuous random variables $X_i s$ with distribution function F and density function f. Then the density function and survival function of U_n, which are denoted by f_{U_n} and $\bar{F}_{U_n}$, respectively, are given by

$$f_{U_n}(x) = \frac{[-log\bar{F}(x)]^{n-1}}{(n-1)!} f(x), \qquad x > 0, n \geq 1,$$

$$\bar{F}_{U_n}(x) = \sum_{j=0}^{n-1} \frac{\left(-log\bar{F}(x)\right)^j}{j!} \quad \bar{F}(x) = \frac{\Gamma\left(n; -log\bar{F}(x)\right)}{\Gamma(n)} \tag{5.3.1}$$

where $\Gamma(a; x)$ is known as the incomplete gamma function and is defined as

$$\Gamma(a; x) = \int_x^\infty u^{a-1} e^{-x} dx, \qquad a, x > 0$$

Arnold et al. [1998].

Remark 5.3.1 The survival function $(5.3.1)$ arises naturally in reliability theory. Assume that a system is put in operation at time $t = 0$. When the system fails, it may be restored to a condition identical to that immediately before failure. That is, the failure rate after repair remains the same as that immediately prior to failure. This kind of repair is called the minimal repair. It is shown that the epoch times of repairs follow a non homogeneous Poisson process with cumulative intensity function $R(t) = -log\bar{F}(t)$. Hence, if $T_1 \leq T_2 \leq \cdots$ denote the epoch times of the repairs then the survival function of $T_n, n = 1,2, \ldots$. is equal to the survival function of record values in $(5.3.1)$ Gupta and Kirmani [1988].

Using the definition of RRE, the result of the following lemma is easy to verify.

Lemma 5.3.1 Let U_n^* denote the nth upper record value from a sequence of observations from $U(0,1)$. Then

$$H_\alpha(U_n^*;t) = \frac{1}{1-\alpha} log \frac{\Gamma\big(\alpha(n-1)+1; -log(1-t)\big)}{\Gamma^\alpha\big(n; -log(1-t)\big)}.$$

The next theorem represents the RRE of upper record U_n in terms of upper record U_n^* of uniform distribution. First, we need the following notation.

Notation: We use the notation $X \sim \Gamma_t(a,\lambda)$ to show that X has a truncated Gamma distribution with density function

$$f(x) = \frac{\lambda^a}{\Gamma(a;t)} x^{a-1} e^{-\lambda x}, \qquad x > t > 0,$$

where $a > 0$ and $\lambda > 0$.

Theorem 5.3.1 The RRE of U_n can be written in terms of the RRE of U_n^* as follows:

$$H_\alpha(U_n;t) = H_\alpha\big(U_n^*; F(t)\big) + \frac{1}{1-\alpha} logE\big[f^{\alpha-1}\big(F^{-1}(1-e^{-V_n})\big)\big] \qquad (5.3.2)$$

where $V_n \sim \Gamma_{-log\bar{F}(t)}(\alpha(n-1)+1,1)$.

Proof The proof follows easily from the definition of RRE of U_n in which one needs to make substitution $u = -log\bar{F}(x)$ in the integrant. Then after some simple algebraic manipulations the result follows.

From Eq. (5.3.2), it is easily seen that the residual Shannon entropy of nth upper record value of an absolutely continuous distribution function F can be written in terms of the residual Shannon entropy of nth upper record value of $U(0,1)$ as follows:

$$H(U_n;t) = H\big(U_n^*; F(t)\big) - E\big[logf\big(F^{-1}(1-e^{-V_n})\big)\big] \qquad (5.3.3)$$

where $V_n \sim \Gamma_{-log\bar{F}(t)}(n,1)$.

Example 5.3.1 Let X have Weibull distribution with density

$$f(x) = \beta\lambda^\beta (x-\mu)^{\beta-1} e^{-[\lambda(x-\mu)]^\beta}, \qquad x \geq \mu.$$

Here, $F^{-1}(x) = \frac{1}{\lambda}\left(-log(1-x)\right)^{\beta} + \mu$. Then we have for $\beta \geq 1$,

$$E\left[f^{\alpha-1}\left(F^{-1}(1-e^{-V_n})\right)\right]$$

$$= \frac{(\lambda\beta)^{\alpha-1}}{\Gamma\left(\alpha(n-1)+1; \left(\lambda(t-\lambda)\right)^{\beta}\right)}\left(\frac{\Gamma\left(\frac{1}{\beta}(1-\alpha)+n\alpha; \alpha\left(\lambda(t-\mu)\right)^{\beta}\right)}{\alpha^{\frac{1}{\beta}(1-\alpha)+n\alpha}}\right).$$

Therefore

$$H_\alpha(U_n;t) = \frac{1}{1-\alpha} log \frac{\Gamma\left(\frac{1}{\beta}(1-\alpha)+n\alpha; \alpha\left(\lambda(t-\mu)\right)^{\beta}\right)}{\Gamma^{\alpha}\left(n; \left(\lambda(t-\mu)\right)^{\beta}\right)} - log(\lambda\beta)$$

$$-\frac{1}{\beta} log\alpha - \frac{n\alpha}{1-\alpha} log\alpha.$$

The following theorem gives a lower bound for RRE of nth record in terms of RRE of uniform distribution. The proof of the theorem is similar to proof of Theorem 5.2.2 and hence is omitted.

Theorem 5.3.2 Suppose that $M = f(m) < \infty$ where m is the mode of X and assume that the assumptions of Theorem 5.3.1 are met. Then for $\alpha > 0$,

$$H_\alpha(U_n;t) \geq H_\alpha\left(U_n^*; F(t)\right) - logM \qquad .$$

It is seen that the bound in Theorem 5.3.2 is very similar to the bound in part (b) of Theorem 5.2.2. This similarity results from assumptions of theorems and representations $(5.2.5)$ and $(5.3.2)$ where "in both of them the corresponding RRE of uniform distribution and density-quantile function appear.

Example 5.3.2 The density functions of the mixture of two Pareto distributions with parameters θ_1 and θ_2 is

$$f(x) = \gamma\theta_1 x^{-\theta_1-1} + (1-\gamma)\theta_2 x^{-\theta_2-1}, \quad x \geq 1, 0 < \gamma < 1,\ \theta_1 > \theta_2 > 0.$$

Since, the mode of this distribution is $m = 1$, we have

$$H_\alpha(U_n;t) \geq \frac{1}{1-\alpha} log \frac{\Gamma\big(\alpha(n-1)+1; -log\bar{F}(t)\big)}{\Gamma^\alpha\big(n; -log\bar{F}(t)\big)}$$

$$-log(\gamma\theta_1 + (1-\gamma)\theta_2),$$

where $-log\bar{F}(t) = (\theta_1+\theta_2)logt - log\big(\gamma t^{\theta_2} + (1-\gamma)t^{\theta_1}\big)$.

The following theorem investigates the monotone behavior of RRE of upper records in terms of n.

Theorem 5.3.3 Let $\{X_i, i \geq 1\}$ be a sequence of $i.i.d$ random variables from distribution function F having an increasing density function f. If $\{U_n, n \geq 1\}$ represents the sequence of upper record values corresponding to F, then $H_\alpha(U_n;t)$ is decreasing in n.

Proof Using Theorem 5.3.1, we have

$$H_\alpha(U_{n+1};t) - H_\alpha(U_n;t)$$

$$= \Delta_\alpha(n;t) + \frac{1}{1-\alpha} log \frac{E\left[f^{\alpha-1}\big(F^{-1}(1-e^{-V_{n+1}})\big)\right]}{E\left[f^{\alpha-1}\big(F^{-1}(1-e^{-V_n})\big)\right]} \qquad (5.3.4)$$

where

$$\Delta_\alpha(n;t) = H_\alpha\big(U^*_{n+1}; F(t)\big) - H_\alpha\big(U^*_n; F(t)\big) \qquad (5.3.5)$$

and

$$V_n \sim \Gamma_{-log\bar{F}(t)}(\alpha(n-1)+1,1).$$

On taking $u(x) = x$ and $v_\alpha(x) = x^{-\alpha}e^{-x}$ and using Lemma 5.2.2, we can show that $H_\alpha\big(U^*_n; F(t)\big)$ is decreasing in n and hence $\Delta_\alpha(n;t) \leq 0$. One can show that $V_n \leq_{lr} V_{n+1}$ and hence $V_n \leq_{st} V_{n+1}$. This implies that, for $0 \;\; \alpha \;\; 1$ ($\alpha \;\; 1$),

$$E\left[f^{\alpha-1}\big(F^{-1}(1-e^{-V_n})\big)\right] \geq (\leq)\, E\left[f^{\alpha-1}\big(F^{-1}(1-e^{-V_{n+1}})\big)\right].$$

Hence, we have for $\alpha \;\; 0, \alpha \neq 1$

$$\frac{1}{1-\alpha} log \frac{E\left[f^{\alpha-1}\left(F^{-1}(1-e^{-V_{n+1}})\right)\right]}{E\left[f^{\alpha-1}\left(F^{-1}(1-e^{-V_n})\right)\right]} \leq 0.$$

Thus, from the last inequality and inequalities $(5.3.4)$ and $(5.3.5)$, we conclude that $H_\alpha(U_{n+1}^*; t) - H_\alpha(U_n^*; t) \leq 0$. This completes the theorem.

An example of the distributions for which this theorem can be applied is the power distribution with distribution function $F(x) = x^\beta,\ 0 < x < 1,\ \beta > 1$.

5.4 Characterization Based on First-Order Statistics

Rao et al. $[2004]$ introduced a new measure of information that extends the Shannon entropy to continuous random variables, and called it cumulative residual entropy (CRE). He show that it is more general than the Shannon entropy and possesses more general mathematical properties than the Shannon entropy. It can easily computed from sample data and its estimation asymptotically converges to the true value. CRE has applications in reliability engineering and computer vision, Rao $[2005]$. This measure is based on the cumulative distribution function $(\mathrm{c.d.f})$ F and is defined as follows:

$$CRE(X) = -\int_0^\infty p(|X| > x) log p(|X| > x) dx.$$

In reliability theory, CRE is based on survival function $\bar{F}(x)$, and is defined as

$$CRE(X) = -\int_0^\infty \bar{F}(x) log \bar{F}(x) dx.$$

In this , we suppose X is a positive continuous random variable. If we use change of variable $u = \bar{F}(x)$, then

$$CRE(X) = -\int_0^1 \frac{u log u}{f(F^{-1}(1-u))} du. \tag{5.4.1}$$

where F^{-1} is the inverse function of F.

Let $X_{1:n}$ be the first-order statistic in a random sample of size n from a positive and continuous random variable X with c.d.f F and p.d.f f, then the c.d.f of $X_{1:n}$ is given by

$$F_{X_{1:n}}(x) = 1 - \bar{F}^n(x)$$

$$CRE(X_{1:n}) = -n\int_0^\infty \bar{F}^n(x)log\bar{F}^n(x)dx$$

By change of variable $u = \bar{F}(x)$, we have

$$CRE(X_{1:n}) = -n\int_0^1 \frac{u^n logu}{f\left(F^{-1}(1-u)\right)}du \qquad (5.4.2)$$

First, let us look at the following examples.

Example 5.4.1 Suppose X has a Pareto (α, β) distribution, with shape parameter $\alpha > 0$ and scale parameter $\beta > 0$. That is, the p.d.f $f(x)$ is given by $f(x) = \frac{\alpha\beta^\alpha}{x^{\alpha+1}}, x \geq \beta$. Using (5.4.1) and (5.4.2), the $CRE(X)$ and $CRE(X_{1:n})$ are as follows:

$$CRE(X) = \frac{\alpha\beta}{(\alpha-1)^2}, \qquad \alpha > 1$$

$$= \infty, \qquad \alpha \leq 1,$$

and

$$CRE(X_{1:n}) = \frac{n\alpha\beta}{(n\alpha-1)^2}, \quad \alpha > \frac{1}{n}$$

$$= \infty, \qquad \alpha \leq \frac{1}{n}$$

Let $\alpha > 1$ and $\Delta_1 = CRE(X) - CRE(X_{1:n})$, then $\Delta \geq 0$, that means for $\alpha > 1$, uncertainty of X is more than $X_{1:n}$. Similarly, this property is obtained for every α and β if we replace $CRE(X)$ and $CRE(X_{1:n})$ by $H(X)$ and $H(X_{1:n})$ respectively. We can also show that for $n > \frac{1}{\alpha}, \Delta_1$ is an increasing function of n.

Example 5.4.2 A non negative random variable X is Weibull distribution, if its c.d.f is

$$F(x) = 1 - \exp(-\lambda^q x^q), \qquad \lambda > 0, q > 0, x > 0$$

where λ and q are, respectively, scale and shape parameters. We can show that $E(X) = \frac{1}{\lambda}\Gamma\left(1 + \frac{1}{q}\right)$ and $CRE(X) = \frac{1}{\lambda q}\Gamma\left(1 + \frac{1}{q}\right)$. Thus, $\frac{CRE(X)}{E(X)} = \frac{1}{q}$. If $q = 1$, then X is standard exponential distributed and this ratio is equal 1.

By (5.4.2), we have

$$CRE(X_{1:n}) = \frac{n}{\lambda q}\int_0^1 u^{n-1}(-logu)^{\frac{1}{q}}du$$

$$= \frac{1}{\lambda q n^{\frac{1}{q}}}\Gamma\left(1 + \frac{1}{q}\right).$$

On the other hand, $E(X_{1:n}) = \frac{1}{\lambda n^{\frac{1}{q}}}\Gamma\left(1 + \frac{1}{q}\right)$. Thus, $\frac{CRE(X_{1:n})}{E(X_{1:n})} = \frac{1}{q}$. This result shows that for all n, in Weibull family, this ratio is constant. If $q = 1$, then for standard exponential distribution this ratio is equal to 1, for all n. We also have

$$\Delta_2 = CRE(X) - CRE(X_{1:n}) = \frac{1}{\lambda q}\Gamma\left(1 + \frac{1}{q}\right)\left(1 - \frac{1}{n^{\frac{1}{q}}}\right) \geq 0,$$

for all n and is increasing in n.

In the following theorem, we show that only in Weibull family the ratio $\frac{CRE(X_{1:n})}{E(X_{1:n})}$ is constant. The following lemma is used in the following theorem. It is known in the literature as Müntz-Szász Theorem, which is often invoked in moment-based characterization theorems, Kamps [1998] and Borwein and

Erdelyi [1995].

Lemma 5.4.1 For any increasing sequence of positive integers $\{n_j, j \geq 1\}$, the sequence of polynomials (X^{n_j}) is complete on $L(0,1)$, if and only if $\sum_{j=1}^{\infty} n_j^{-1} = \infty$.

Theorem 5.4.1 Suppose that $X_1, X_2, \ldots, X_n$ are positive, independent and identically distributed $(\mathrm{i.i.d})$ observations from an absolutely continuous $\mathrm{c.d.f}$ $F(x)$ and $\mathrm{p.d.f}$ $f(x)$. Then F belongs to Weibull family, if and only if $\frac{CRE(X_{1:n})}{E(X_{1:n})} = c\ (> 0)$, for all $n = n_j, j \geq 1$, such that $\sum_{j=1}^{\infty} n_j^{-1} = \infty$.

Proof By Example 5.4.1, necessity is trivial, hence it remains to prove the sufficiency part. By using change of variable $\bar{F}(x) = u$ in $E(X_{1:n}) = \int_0^{\infty} nxf(x)\bar{F}^{n-1}(x)dx$, we have

$$E(X_{1:n}) = n\int_0^1 F^{-1}(1-u)u^{n-1}du. \qquad (5.4.3)$$

Using (5.4.2) and (5.4.3), we have

$$\frac{CRE(X_{1:n})}{E(X_{1:n})} = -\frac{\int_0^1 \frac{u^n logu}{f\left(F^{-1}(1-u)\right)}du}{\int_0^1 F^{-1}(1-u)u^{n-1}du} \qquad (5.4.4)$$

If (5.4.4) coincides c, we can conclude that

$$\int_0^1 u^{n-1}\left[\frac{ulogu}{f\left(F^{-1}(1-u)\right)} + cF^{-1}(1-u)\right]du = 0 \qquad (5.4.5)$$

If (5.4.5) holds for $n = n_j, j \geq 1$, such that $\sum_{j=1}^{\infty} n_j^{-1} = \infty$, then from Lemma 5.4.1, we have

$$\frac{(1-v)log(1-v)}{f\left(F^{-1}(v)\right)} + cF^{-1}(v) = 0 \qquad a.e, v \in (0,1).$$

Since $\frac{d}{dv}F^{-1}(v) = \frac{1}{f\left(F^{-1}(v)\right)}$, it then follows:

$$(1-v)log(1-v)\frac{d}{dv}F^{-1}(v)+cF^{-1}(v)=0 \quad a.e, v\,\epsilon\,(0,1).$$

After solving this differential equation, we can result that $F^{-1}(v)=c_1[-log(1-v)]^c, v\,\epsilon\,(0,1)$, thus $F(x)=1-exp(-(\frac{x}{c_1})^{\frac{1}{c}}), x>0$. This means that F belongs to the Weibull family.

Theorem 5.4.2 Let X and Y be two positive random variable with $\mathrm{p.d.f's}\, f(x)$ and $g(x)$ and absolutely continuous $\mathrm{c.d.f's}\, F(x)$ and $G(x)$, respectively. Then F and G belong to the same family of distributions, but for a change in location, if and only if

$CRE(X_{1:n})=CRE(Y_{1:n})$

for $n=n_j, j\geq 1$ such that $\sum_{j=1}^{\infty}n_j^{-1}$ is infinite.

Proof The necessity is trivial, hence it remains to prove the sufficiency part. By (5.4.2), if $CRE(X_{1:n})=CRE(Y_{1:n})$, then we have

$$\int_0^1 u^n logu\left[\frac{1}{f\big(F^{-1}(1-u)\big)}-\frac{1}{g\big(G^{-1}(1-u)\big)}\right]du=0. \quad (5.4.6)$$

If (5.4.6) holds for $n=n_j, j\geq 1$, such that $\sum_{j=1}^{\infty}n_j^{-1}=\infty$, then from Lemma 5.4.1 we can conclude that $f\big(F^{-1}(t)\big)=g\big(G^{-1}(t)\big), 0<t<1$. Since $\frac{d}{dv}F^{-1}(t)=\frac{1}{f\left(F^{-1}(t)\right)}$, we have $\frac{d}{dv}F^{-1}(t)=\frac{d}{dv}G^{-1}(t), 0<t<1$. It then follows that $F^{-1}(t)=G^{-1}(t)+d, 0<t<1$. This means F and G belong to the same family of distributions, but for a location shift.

Bibliography

Abraham, B. and Sankaran, P. G., (2006), "Renyi's entropy for residual lifetime distribution." Statistical Papers, Springer- Verlag, Vol. 46, PP. 17-30.

Arimoto, S., (1971), "Information theoretical considerations on estimation theory." Information and Control, Vol. 19, PP. 181-190.

Arnold, B. C., Balakrishnan, N., Nagaraja, H. N., (1998), "Records." Wiley, New York.

Arnold, Balakrishanan, N., Nagaraja, H. N., (1992), "First course in order statistics." Wiley, New York.

Asadi, M. and Ebrahimi, N., (2000), "Residual entropy and its characterizations in terms of hazard function and mean residual life function." Statist. Prob. Lett., Vol. 49, 263-269.

Asadi, M., Ebrahimi, N. and Soofi, E., (2005), "Dynamic generalized information measures." Stat. Prob. Letters, Vol. 71, PP. 85-98.

Asadi, M., Zohrevand, Y., (2007), "On the dynamic cumulative residual entropy." Journal of Statistical Planning and Inference, Vol. 137, PP. 1931-1941.

Ash, R. B., (1990), "Information theory." Dover, New York.

Azlarov, T. and Volodin, N., (1986), "Characterization problems associated with experimental distribution." Springer-Verlag, Berlin.

Baig, M. A. K. and Javid (2008), "Some results on survival exponential entropies." International Journal of Pure and Applied Mathematics, Vol. 24, PP. 91-98.

Baig, M. A. K. and Javid (2008), "Survival and generalized survival exponential entropies for Parto I and II kind distribution." Antarctica Journal of Mathematics, Vol. 5 (1), PP. 87-96.

Baig, M. A. K. and Javid, (2008), "Generalized past entropy for some lifetime models." Presented in 4th J&K science congress held in University of Kashmir, Srinagar.

Baig, M. A. K. and Javid, (2008), “Generalized residual entropy function and its applications.” European Journal of Pure and Applied Mathematics, Vol. 1 (4), PP. 30-40.

Baig, M. A. K. and Javid, (2010), “Renyi’s entropy for past lifetime distributions.” International Journal of Pure and Applied Mathematics (appear).

Baig, M. A. K. and Javid, (2010), “Measures of discrimination in survival analysis.” (Communicated).

Baig, M. A. K. and Javid., (2008), “Characterization of generalized residual entropy function ” Presented in International conference on new trends in statistics and optimization held in university of Kashmir, Srinagar.

Baig, M. A. K. and Javid., (2009), “Generalized information measure and its properties.” Journal of Inequalities of Pure and Applied Mathematics, Vol. 10,

Baig, M. A. K. and Javid., (2009), “Some new results on survival exponential entropies.” Bulletin of the Allahabad Mathematical Society, Vol. 24 (1), PP. 91-98.

Baig, M. A. K., Javid. and Ibrahim, M., (2009), “Generalized past entropy in survival analysis.” Global Journal of Pure and Applied Mathematics, Vol. 5 (3), PP. 229-237.

Baig, M. A. K., Javid and Jan, T. R., (2009), “Past residual entropy expressions for some lifetime distributions.” International Journal of Contemporary Mathematical Science (appear).

Balasubrahmanyan, P. and Siromoney, G., (1968), “A note on entropy of Telgu prose.” Information and Control, Vol. 13, PP. 281-285.

Bar- Hillel, Y., (1964), “Language and information.” Addision Wasley Pub. Co., The Jerusalum Academic Press Limited.

Baratpour, S., Ahmadi, J., Arghami, N. R., (2007), “Some characterization based on entropy of order statistics and record values.” Commu. Statist. Theor. Meth. Vol. 36, PP. 47-57.

Baratpour, S., Ahmadi, J., Arghami, N. R., (2007), “Entropy properties of record statistics.” Statistical Papers, Vol. 48, PP. 197-213.

Baratpour, S., Ahmadi, J., Arghami, N. R., (2008), “Some characterization based on Renyi entropy of order statistics and record values.” Journal Statist. Plann. Infer. Vol. 138, PP. 2544-2551.

Baratpour, S., Ahmadi, J., Arghami, N. R., (2008), “Characterizations based on renyi entropy of order statistics and record values.” Journal of Statistical Planning and Inference, Vol. 138, PP. 2544-2551.

Barlow, R. and Proschan, F., (1981), “Statistical theory of reliability and life testing: Probability models.” Silver Spring, MD.

Belzunce, F., Navarro, j., Ruiz, J.M. and Aguila, Y., (2004), “Some results on residual entropy function.” Metrika, Vol. 59, PP. 147-161.

Bickel, P. j. and Doksum, K. A., (1969), “Test for monotone failure rate based on normalized spacing.” Ann. Math. Statist, Vol. 40, PP. 1216-1235.

Block, H. W., Borges, W. S., Savits, T. H., (1985), “Age dependent minimal repair.” Journal of Applied Probability, Vol. 22, PP. 370-385.

Block, H.W., Savits, T. H. and Sing, H., (1998), “The reversed hazard rate function.” Prob. Eng. Inf. Sci, Vol. 12, PP. 69-90.

Boekee, E. and Van Der Lubbe, J. C. A., (1980), “The R-norm information measure.” Information and Control, Vol. 45, PP. 136-155.

Borwein, P., Erdelyi, T., (1995), “Polynomials and polynomials inequalities.” Springer, New York.

Brillouin, L., (1956), “Science and Information theory.” New York, Academic Press.

Bryson, M. C. and Siddiqui., M. M., (1969), “Some criteria for aging.” JASA, Vol. 64, PP. 1472-1484.

Campbell, L. L., (1966), “Exponential entropy as a measure of extent of a distribution.” Zeitschrift. Vol. 5, PP. 217-225.

Chandra, N. K. and Roy, D., (2001), "Some results on reversed hazard rate." Prob. Eng. Inf. Sci., Vol. 15, PP. 95-102.

Chen, C. H., (1973), "Statistical pattern recognition." Recelle Park, New York, Hoyderc Book Co.

Courant, R. and John, F., (1989), "Introduction to Calculus and Analysis." Vol. II, Springer-Verlag, New York.

Cover, T. M. and Thomas, J., (1991), "Elements of information theory." Willey, New York.

Cox, D. R. and Oakes, D., (1984), "Analysis of survival data." Chapman and Hall, New York.

Cox, D. R., (1972), "Regression models and life tables." Journal of Royal Statistical Society Ser. B, Vol. 34, PP. 187-202.

Crescenzo, A. D. and Longobardi, M., (2002), "Entropy based measure of uncertainty in past lifetime distribution." Journal of Applied Probability, Vol. 39, PP. 434-440.

Crescenzo, A. D. and Longobardi, M., (2004), "A measure of discrimination between past lifetime distributions." Statistics and Probability Letters, Vol. 67, PP. 173-182.

Crescenzo, A. D. and Longobardi, M., (2006), "On weighted residual and past entropies." Sci. Math. Jpn. Vol. 64, PP. 255-266.

David, H. A., Nagaraja, N. H., (2003), "Order statistics." Wiley, New York, Vol. 3.

Drissi, N., Chonavel, T. and Boucher, J. M., (2008), "Generalized cumulative residual entropy for distributions with unrestricted supports." Research Letters in Signal Processing. Article Id 790607.

Ebrahimi, N. and Habibullah M., (1990), "Testing whether the survival function is NBU." Biometrika, Vol. 77, PP. 212-215.

Ebrahimi, N. and Kirmani, S. N. U. A., (1996), "A characterization of the proportional hazard model through a measure of discrimination between two residual life distributions." Biometrika, Vol. 83, PP. 233-235.

Ebrahimi, N. and Kirmani, S. N. U. A., (1996), “A measure of discrimination between two residual life-time distributions and its applications.” Ann. Inst. Statist. Math., Vol. 48, PP. 257-265.

Ebrahimi, N. and Kirmani, S.N.U.A., (1996), “Some results on ordering of survival functions through uncertainty.” Statist. Probability. Lett., Vol. 29, PP. 167-176.

Ebrahimi, N. and Pellerey, F., (1995), “New partial ordering of survival functions based on notion of uncertainty.” Journal of Applied Probability, Vol. 32, PP. 202-211.

Ebrahimi, N. and Zahedi, H., (1992), “Memory ordering of survival functions.” Statistics Vol. 23, PP. 337-345.

Ebrahimi, N., Habibullah M. and Soofi, S., (1992), “Testing exponentiality based on Kullback Leibler information.” JRSS Ser. B, Vol. 54, PP. 739-748.

Ebrahimi, N., (1996), “How to measure uncertainty in the residual life distribution.” Sankhya Ser. A, Vol. 58, PP. 48-56.

Ebrahimi, N., (1997), “Testing whether lifetime distribution is decreasing uncertainty.” Journal of Statistical Planning and Inference, Vol. 64, PP. 919.

Ebrahimi, N., (2000), “The maximum entropy method for lifetime distribution.” Sankhya A, Vol. 62, PP. 236-243.

Ebrahimi, N., Habibullah, M. and Soofi, E. S., (1993), “The role of information theory in reliability analysis.” In advance in Reliability, A. P. Basu, Ed. Amsterdam, The Netherlands: Elsevier, PP. 89-103.

Ebrahimi, N., Soofi, S. E., Zahedi, H., (2004), “Information properties of order statistics and spacings.” IEEE Transaction on Information Theory IT 50, PP. 177-183.

Galambos, J., (1978), “The asymptotic theory of extreme order statistics.” Wiley, New York.

Ghitany, M. E., El-Saidi, M. A. and Khalil., Z., (1995), “ Characterization of a general class of life testing models.” Journal of Applied Probability, Vol. 32, PP. 548-553.

Gupta, R. C. and Kirmani S. N. U. A., (1988), “Closure and monotonicity properties of nonhomogeneous Poisson processes and record values.” Probability in the Engineering and Informational Sciences, Vol. 2, PP. 475-484.]

Gupta, R. C. and Kirmani, S. N. U. A., (1990), “The role of weighted distribution in stochastic modeling.” Comm. Statist. Theory Methods, Vol. 19, PP. 3147-3162.

Gupta, R. D. and Nanda A. K., (2002), “α and β entropies and relative entropies of distributions.” Journal of Statist. Theory Application, Vol. 3, PP. 177-190.

Hall, P. and Morton, S. C., (1993), “ On the estimation of entropy.” Ann. Inst. Statist. Math., Vol. 45, PP. 69-88.

Hall, W. J., Wellner, J. A., (1981), “Mean residual life, in: Csorgo, M. Dawson, Rao, J. N. K., Saleh, A. K. Md. E.” Statistics and Related Topics, PP. 169-184.

Hartley, R. V. L., (1928), “Transmission of information.” Bell. Sys. Tech. Journal, Vol. 7, PP. 535-563.

Havrda, J. and Charvat, F., (1967), “Quantification method of classification process: concept of structural α entropy.” Kybernetika, Vol. 3, PP. 30-35.

Hooda, D. S. and Kumar, P., (2007), “Generalized residual entropies in survival analysis.” Journal of Applied Probability and statistics, Vol. 2, PP. 241-249.

Jain, K., Singh, H. and Bagai, I., (1989), “Relations for reliability measures of weighted distributions.” Comm. Statist. Theory Methods, Vol. 18, PP. 4393-4412.

Jaynes, E. T., (1957), “Information theory and statistical mechanics.” Physical Review, Vol. 106, PP. 620-630.

Jeevanand, E. S. and Sathar, E. I., (2009), “Estimation of residual entropy function for exponential distribution from censored sample.” ProbStat Forum, Vol. 2, PP. 68-77.

Johnson, N. L. and Kotz, S., (1970), “Distribution in statistics, continuous univariate distributions.” Wiley, New York.

Jumarie, G., (19990, “Relative information.” Springer, PP. 55-57.

Kamps, U., (1995), “A concept of generalized order statistics.” Teubner, Stuttgart.

Kamps, U., (1998), “Characterizations of distributions by recurrence relations and identities for moments of order statistics.”

Kapur, J. N., (1967), “Generalized entropy of order α and type β.” The Mathematical Seminar (Delhi), Vol. 4, PP. 78-94.

Kapur, J. N., (1968), “On some applications of dynamic programming to information theory.” Proceedings 2nd Academic Science, Vol. 68, I-II.

Keilson, J. and Sumita, U., (1982), “Uniform stochastic ordering and related inequalities.” Canadian Journal of Statistics, Vol. 10, PP. 181-198.

Kerridge, D. F., (1961), “Inaccuracy and inference.” Journal of Royal Statistical Society, Series B, Vol. 23, PP. 184-194.

Khinchin, A. J., (1957), “Mathematical foundation of information theory.” Dover, New-York..

Kochar, S. C., (1990), “Some partial ordering results values.” Comm. Statist. Theory Methods, Vol. 19, PP. 299-306.

Kotlarski, I. I., (1972), “On a characterization of some probability distributions by conditional expectation.” Sankhya A, Vol. 34, PP. 461-466.

Kotz, S. and Shanbhag, D. N., (1980), “Some new approaches to probability distributions.” Adv. Applied Probability, Vol. 12, PP. 903-921.

Kullback, S. and Leibler, R. A., (1951), “On information and sufficiency.” Ann. Math. Statist., Vol. 22, PP. 79-86.

Kullback, S., (1959), “Information theory and statistics.” Wiley, New York.

Kumar, V., Taneja, H. C., Srivasta, R., (2010), “On dynamic Renyi cumulative residual entropy measure.” Journal of Statistical Theory and Applications.

Kurths, J., et al., (1995), “Quantitative analysis of heart rate variability.” Chaos, Vol. 1, PP. 88-94.

Lazo, A. C. G. V. and Rathie, P. N., (1978), “On the entropy of the continuous probability distributions.” IEEE Trans. Inform. Theory, Vol. 24, PP. 120-122.

Marshall, A. W. and Olkin, I., (2007), “Life distributions. Structure of nonparametric, semiparametric and parametric families.” Springer Series in Statistics. Springer, New York.

Mei, M., (1978), “The theory of genetic distance and evaluation of human race.” Japan Journal of Human Genetics, Vol. 23, PP. 341-369.

Morales, D., Pardo, L. and Vajda, I., (1997), “Some new statistics for testing hypothesis in parametric models.” Journal of Multivariate Analysis, Vol. 62, PP. 137-168.

Mukerjee, S. P. and Islam, A., (1983), “A finite range distribution of failure times.” Navel, Res., Logist, Vol. 30, PP. 487-491.

Muth, E. J., (1980), “Memory as a property of probability distribution.” IEEE Tran. in Reliability, Vol. 29, PP. 160-165.

Nadarajah, S. and Zografos, K., (2003), “Formulas for Renyi information and related measures for univariate distributions.” Information Science, Vol. 155, PP. 119-138.

Nadarajah, S. and Zografos, K., (2005), “Expression for Renyi and Shannon entropies for bivariate distribution.” Information Science, Vol. 170, PP. 173-189.

Nagaraja, H. N., (1990), “Some reliability properties of order statistics.” Comm. Statist. Theory Methods 19, 307-316.

Nair, K. R. M. and Rajesh, G., (1998), “Characterization of probability distributions using the residual entropy function.” Journal of the Indian Statistical Association. Vol. 36, PP. 157-166.

Nanda, A. K. and Gupta, R. D., (2001), “Some properties of reversed hazard rate function.” Statistical Methods, Vol. 3, PP. 108-124.

Nanda, A. K. and Paul, P., (2005), "Some properties of past entropy and their applications ." Metrika, Vol. 64, PP 47-61.

Nanda, A. K. and Paul, P., (2006), "Some results on generalized past residual entropy." Journal of Statistical Planning and Inference, Vol. 136, PP. 3659-3674.

Nanda, A. K. and Paul, P., (2006), "Some results on generalized residual entropy." Information Science, Vol. 176, PP. 24-47.

Nanda, A. K. and Shaked, M., (2001), "The hazard rate and the reversed hazard rate orders with applications to order statistics." Ann. Inst. Statist. Math., Vol. 53, PP. 853-864.

Nanda, A. K., Jain, K., (1999), "Some weighted distributions results on univariate and bivariate cases." Journal of Statistical Planning Inference, Vol. 77, PP. 169-180.

Nanda, A. K., Singh, H., Misra, N., Paul, P., (2003), "Reliability properties of reversed residual lifetime." Communications in Statistics: Theory and Methods, Vol. 32, PP. 2031-2042.

Navarro, J. and Ruiz, J. M., (2004), "Characterization from relationships between failure rate functions and conditional moments." Commun. Statist. Theor. Meth., Vol. 33, PP. 3159-3171.

Navarro, J., Belzunce, F., Ruiz, J. M. and Del Aguila, Y., (2002), "Some results on residual entropy function." 3rd Internal. Conf. Math. Methods Reliability Trondheim Norway, PP. 17-20.

Navarro, J., Ruiz, J. M. and del Aguila, Y., (2001), "Parametric estimation from weighted samples." Biometrical Journal, Vol. 43, PP. 297-311.

Navarro, J., Ruiz, J. M. and del Aguila, Y., (2003), "How to detect baised samples?" Biometrical Journal, Vol. 45, PP. 91-112.

Navarro, J., Ruiz, J. M. and del Aguila, Y., (2006), "Multivariate weighted distributions—A review and some extentions." Statistics, Vol. 40, PP. 51-64.

Navarro, J., Ruiz, J. M. and del Aguila, Y., (2008), "Characterizations and ordering properties based on log-odds functions." Statistics, PP. 313-328.

Nyquist, H., (1924), "Certain factors affecting telegraphy speed." Bell. System Tech. Journal Vol. 3 PP. 324-346.

Nyquist, H., (1928), "Certain topics in telegraphy transmission theory." AIEE Trans., Vol. 47, PP. 617.

Oluyede, B. O., (1999), "On inequalities and selection of experiments for length biased distributions." Prob. Eng. Inf. Sci. Vol. 13, PP.169-185.

Ord, J. K., Patil, G. P., Taillie, C., (1981), "The choice of a distribution to describe personal incomes." Dordrecht-Boston, Massachusetts, PP. 193-201.

Ord, J. K., Patil, G. P., Taillie, C., (1983), "Truncated distributions and measures of income inequality." Sankhya Ser. B, Vol. 45, PP. 413-430.

Park, S., (1995), "The entropy of consecutive order statistics." IEEE Transaction on Information Theory IT 41, PP. 2003-2007.

Patil, G. P. and Rao, C. R., (1977), "The weighted distributions: a survey of their applications." Applications of Statistics, North-Holland Publishing Co., Amsterdam, PP. 383-408.

Quaster, H., (1953), "Information theory and Biology." University of Illinois Press, Urbana.

Quaster, H., (1954), "Information theory and Psychology." Free Press, Glencoe, Illinois, (Proceeding of Conference).

Rajesh, G. and Nair, K. R. M., (1998), "Residual entropy function in discrete time." Far East J. Theo. Stat., Vol. 2, PP. 1-10.

Rajesh, G. and Nair, K. R. M., (2000), "Residual entropy of conditional distributions." Statist. Meth., Vol. 2, PP. 72-80.

Rao, C. R., (1965), "Linear statistical inference and its application." John Wiley and Sons, Inc. New York .

Rao, C. R., (1965), “On discrete distributions arising out of methods of ascertainment.” Sankhya Ser. A, Vol. 27, PP. 311-324.

Rao, C. R., (1982), “Diversity and dissimilarity coefficient: a unified approach.” Theoretical Population Biology, Vol. 21, PP. 24-43.

Rao, M., (2005), “More on a new concept of entropy and information.” Journal Theoret. Probab. Vol. 18, PP. 967-981.

Rao, M., (2005), “More on a new concept of entropy and information.” Journal of Theoretical Probability, Vol. 18, PP. 967-981.

Rao, M., Chen, Y., Vemuri, B. C and Wang, F., (2004), “Cumulative residual entropy: A new measure of information.” IEEE Trans. Inf. Theory, Vol. 50, PP. 1220-1227.

Rathie, P. N., (1972), “Generalized entropy in coding theory.” Metrika, Vol. 18, PP. 216-219.

Renyi, A., (1959), “On the dimension and entropy of probability distribution.” Acla. Math. Acad. Sci. Hung, Vol. 10, PP. 193-215.

Renyi, A., (1961), “On measure of entropy and information.” Proceeding of the fourth Berkeley Symposium on Math. Statist. Prob. University of California Press, Berkelcy, Vol. 1, PP. 547-561.

Rudin, W., (1987), “Real and complex analysis.” McGraw-Hill, PP. 62.

Salomon, D., (1998), “Data compression.” Springer.

Sankaran, P. G. and Gleeja, V. L., (2006), “On bivariae reversed hazard rates.” J. Japan Statist. Soc. Vol. 36, PP. 213.224.

Sankaran, P. G. and Gupta, R. P., (1999), “Characterization of lifetime distributions using measure of uncertainty.” Calcutta Statistical Association Bulletin Vol. 49, PP. 159-166.

Sen, A., (1973), “On economic inequality.” Oxford University Press, London.

Shaked, M. and Shanthikumar, J. G., (1994), “Stochastic orders and their applications.” Boston, Academic Press.

Shaked, M., Shanthikumar, J. G., (2007), “Stochastics orders.” Springer, New York.

Shannon, C.E., (1948), “A mathematical theory of communication.” Bell System Tech. Journal, Vol. 27, PP. 379-423 and 623-656.

Sharma, B. D. and Mittal, P., (1977), “New non –additive measures of relative information.” Journal of Combin. Inform. System Sci., Vol. 2, PP. 122-133.

Sharma, B. D. and Taneja, I. J., (1975), “Entropy of type (α, β) and other generalized measures in information theory.” Metrika, Vol. 22, PP. 205-215.

Sharma, B. D. and Taneja, I. J., (1977), “Three generalized additive measures of entropy.” Elektronische Informations Verarbeitung and Kybernetik Eik., Vol. 13, PP. 419-433.

Sharpe, W. F., (1985), “Investments.” Prentice Hall.

Shore, J. E. and Johxson, R. W., (1980), “Axiomatic derivation of the principle of maximum entropy and the principle of the minimum cross entropy.” IEEE Trans. Inform. Theory, Vol. 26, PP. 26-37.

Singh, H., (1989), “On partial ordering of life time distributions.” Naval Res. Logist. Vol. 36, PP. 103-110.

Song, K., (2001), “Renyi information, log likelihood and an intrinsic distribution measure.” Journal of Statistical Planning and Inference, Vol. 93, PP. 51-69.

Soofi, E. S., (2000), “Principal information theoretic approach.” J. Amer. Statist. Assoc., Vol. 95, PP. 1349-1353.

Sunoj, S. M. and Linu, M. N., (2010), “Dynamic cumulative residual Renyi’s entropy.” Statistics, PP. 1-16.

Sunoj, S. M., Sankaran, P. G. and Maya, S. S., (2009), “Characterizations of the life distributions using conditional expectations of doubly (interval) truncated random variables.” Commun. Statist. Theory and Methods, Vol. 38, PP. 1441-1452.

Takahashi, K., (1988), “A note on hazard rates of order statistics.” Comm. Statist. Theory Methods, Vol, 17(12), PP. 4133-4136.

Taneja, H. C., Kumar, V. and Srivastava, R., (2009), "A dynamic measure of inaccuracy between two residual lifetime distributions." International Mathematical Forum, Vol. 4, PP. 1213-1220.

Taneja, I. J., (1975), "A study of generalized measures in information theory." Ph.D Thesis, University of Delhi.

Taneja, I. J., (1990), "On generalized entropy with applications." In Lectures in Applied Mathematics and Informatics, ed. L. M.. Ricciardi, Manchester University Press, PP. 107-169.

Theil, H., (1967), "Economics and information theory." North Holland, Publishing Company, Amsterdam.

Vajda, I., (1989), "Theory of statistical inference and information." Norwell, MA, Kluwer.

Varma, R. S., (1966), "Generalizations of Renyi's entropy of order α." Journal of Mathematical Science, vol. 1, PP. 34-48.

Wang, F., Vemuri, B. C., Rao, M. and Chen, Y., (2003), "Cumulative residual entropy: A new measure of information and its application to image alignment." In Proc. 9^{th}. IEEE Int. Conf. Computer Vision (ICCV; 03), PP. 548-553.

Wiener, N., (1961), "Cybernetics." The MIT Press, John Willey and Sons Inc., New York, London.

Wong, K. M., Chen, S., (1990), "The entropy of ordered sequences and other order statistics." IEEE Transaction on Information Theory IT 36, PP. 276-284.

Xie, M., Lai, C. D., (1996), "On the increase of the lifetime redundancy." Asia-Pacific Journal of Operational Research, Vol. 13 (2), PP. 171-179.

Zack, S., (1992) "Introduction to reliability analysis." Springer, New- York.

Zadeh, L. A., (1996), "Fuzzy sets." Information and Control, Vol. 8, PP. 94-102.

Zaheeruddin, (1987), "A basic theorem in information theory." The Aligarh Journal of Statistics, Vol. 7, PP. 1-10.